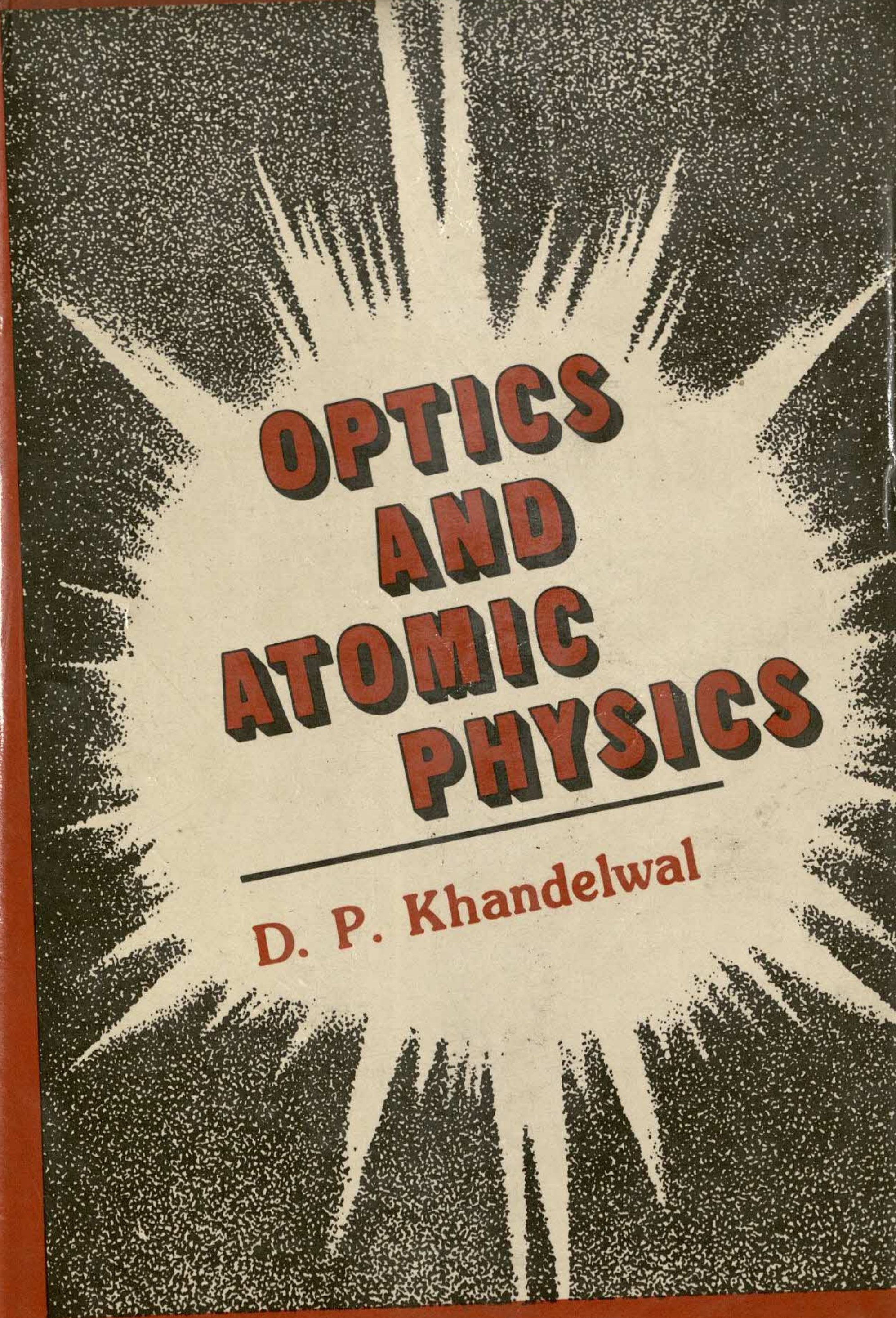


OPTICS AND ATOMIC PHYSICS

D. P. Khandelwal

Himalaya Publishing House

Rs. 150

୨୫୭୯

TEXT BOOK OF
OPTICS AND ATOMIC PHYSICS

UBSPD
UBS Publishers' Distributors Ltd
8/1-B, Chowringhee Lane, Calcutta-700 016
Ph:212451, 249473

TEXT BOOK OF
OPTICS AND ATOMIC PHYSICS

D.P. KHANDELWAL,
M.Sc., PhD.
Formerly Professor of Physics, Agra College Agra
and H.B. Technological Institute, KANPUR

Himalaya Publishing House
BOMBAY • NAGPUR • DELHI

© AUTHOR

Published by: Mrs. Meena Pandey *for*
HIMALAYA PUBLISHING HOUSE
Head Office
"Ramdoot", Dr. Bhalerao Marg
Girgaon (Kelewadi)
Bombay-400 004.
Phones: 363863 360170

Branch Office
"Pooja Apartment" 4-B Murari Lal Street
Ansari Road, Daryaganj
New Delhi-110 002
Phone: 270392

Eighth Revised and Enlarged Edition : 1989

Acc no-16769

Printed at :
Sanjeev Printers
Baberpur, Delhi-110032.

PREFACE TO THE EIGHTH EDITION

The objective of this book has been to provide a coverage on Optics as such and to use the interaction of light with matter as a mother source for atomic physics and related theories. In the past two decades this combine has been found exciting to the students and has been adopted in several Universities. This book has served the students and teachers well in that direction.

In this Edition we have added a full Chapter on *Visual Photometry*, primarily because we find that the terminology used here has not been absorbed well in most available text books. We would suggest that numerical problems be solved in depth to have an understanding of the inter-relations of the different quantities used.

This book has been evolving with successive Editions to keep pace with the developments in the field. We introduced the complex representation of oscillations and waves in the last edition. We have now taken it farther to introduce *Fourier Transforms* in some more details, with a few applications. The relevant mathematics is done at the corresponding level by the student, and this technique is finding increasing use in applied Optics.

With lasers, we now add *Non-linear Optics*. It is time we tell the student that superposition of waves is a limited concept, and nonlinear responses lead to entirely new results and also useful applications. We have introduced one full article on this.

Obviously, to keep the total volume within reasonable size, we have cut down details here and there. But the emphasis on solved examples on a wide range, critical discussion of the assumptions and results, and numerical estimates of what a theory means, has been reatained. The inherent assumption is that a teacher is conducting the course with the students, and there is no need to include avoidable details.

The author will welcome suggestions from the acdemic community for further improvement of the book.

D.P. Khandelwal

Kanpur

December 6, 1988

PREFACE TO THE FIRST EDITION

This book covers the course in Geometrical and Physical Optics for most of the Universities in India. Elements of quantum theory are also introduced from the experiment approach—the photo-electric effect, Compton effect and the series relations in atomic spectra. Special care has been taken to discount the idea of a 'medium' like ether and to say that light waves are made up of oscillating electric and magnetic vectors. However, electro-magnetic theory and relativity as such have not been introduced.

The language of the book has been kept as simple as could be consistent with precision and brevity.

The mathematical steps in the text have on purpose been kept fast, partly to reduce the volume of the book, and partly to allow for different approaches by the individual students. Further, the emphasis must be on the physical assumptions underlying a procedure, rather than the mathematical procedures of integration, summation, etc.

Similarly, detailed descriptions of experiments are also not given. Textbooks in practical physics are available for the purpose. Essential features only are described. There are several instruments in applied optics—geometrical as well as wave-optics. The temptation to include some of them was great. But it was necessary to keep the volume handy and within reasonably low cost. Once the basic instruments are understood well, it should be easy for any teacher to introduce the other instruments. This explains the exclusion of a larger variety of applied instruments. The emphasis has been on basic phenomena and treatment of fewer phenomena in depth and multiple ways.

The inherent harmony in the theory of co-axial image-forming systems is not realised in many texts. In the present text, special care has been taken to emphasise this. For example, the thick lens can be treated as made up of two thin lenses separated by a medium. Again, repeated references have been made to certain universal relations by proving them in several particular cases also. Nothing is better than facing a result repeatedly and then appreciating that it may be of universal application.

In the theory of interference and diffraction, cross-references for proofs have been avoided. It is good to go through the drill repeatedly. Where alternative but equally powerful procedures are available, they have been given separately.

It has been extensively recognised that without a proper appreciation of 'magnitudes', all physics goes flat with the students. The orders of magnitude of dispersion, chromatic aberration, astigmatic images etc., or the orders of magnitude of diffraction spreading, the zone plate, the half-period zones, the fringe-width, the intensity variation etc., must all be calculated in order that the phenomena are properly appreciated. This approach has been uniformly adopted in this text, and is further encouraged through the problems. A close approximation to intensity distribution in Fresnel fringes due to a straight edge is given probably for the first time in this text book.

The problems form an integral part of the text. It is now the standard recognised practice in all advanced teaching methods that the problems must not be limited to reproduction of some parts of the text, but must lead to projection of the student's mind beyond the subject-matter of the text. Recent examination trends are also in that direction. Therefore, we have not simply reproduced examination questions from different universities, but have built up a structure of problems which have a great teaching value in themselves.

It is sincerely hoped that fellow teachers will find this text book exciting and the students will find it interesting and useful.

Agra College, Agra
May 4, 1967 **D. P. Khandelwal**

CONTENTS

1. **Fermat's Principle and Ray Optics** 1-12
    - 1·1 A Close Examination of the Concept of Rays of Light
    - 1·2 Fermat's Principle
    - 1·3 Laws of Reflection from Fermat's Principle
    - 1·4 Laws of Refraction from Fermat's Principle
    - 1·5 Extremum Path ; not Least Path
    - 1·6 Case of a Virtual Point on the path

2. **Cardinal Points of an Optical System** 13-24
    - 2·1 Convention of Signs
    - 2·2 Gauss's General Theory of Image Formation in a Coaxial Symmetrical System
    - 2·3 Conjugate Relationships
    - 2·4 Cardinal Points of an Optical System
    - 2·5 Actual Systems
    - 2·6 Lagrange's and Helmholtz's Laws
    - 2·7 Relationship between f and f' for any Optical System
    - 2·8 Relations between Lateral, Axial and Angular Magnifications

3. **Thick Lens and Lens Combinations** 25-35
    - 3·1 Thick Lens in Air
    - 3·2 Combination of Two Thin Lenses
    - 3·3 Some Special Cases of Lens Combinations
    - 3·4 General Combination of Two Optical Systems or of Two Thick Lenses
    - 3·5 Deviation Method for Locating the Cardinal Points

4. **Dispersion of Light and Chromatic Aberration** 36-44
    - 4·1 Dispersion
    - 4·2 Longitudinal Chromatic Aberration of an Optical System
    - 4·3 Lateral Chromatic Aberration
    - 4·4 Achromatic Combination of Thin Lenses in Contact

4·5 Achromatic Combination of Two Separated Lenses
4·6 Achromatism of Size, not of Position of Images

5. Monochromatic Aberration of Images 45-61
5·1 General Description of Different Aberrations
5·2 Formulae for Spherical Aberration in some cases
5·3 A General Principle for Reducing Spherical Aberration
5·4 Aplanatic Surfaces and Aplanatic Points
5·5 Reduction or Elimination of Spherical Aberration
5·6 Elimination of Coma. Abbe's Sine Condition
5·7 Elimination of Astigmatism
5·8 Elimination of Curvature of the Field
5·9 Reduction or Elimination of Distortion

6. Optical Instruments 62-79
6·1 Entrance Pupil and Exit Pupil
6·2 Field of View
6·3 Need for a Multiple-Lens Eyepiece
6·4 Huygens' Eyepiece
6·5 Ramsden's Eyepiece
6·6 Gaussian Eyepiece
6·7 Kellner's Eyepiece
6·8 Comparison of Huygens', Ramsden's and Kellner's Eyepieces
6·9 Compound Microscope
6·10 Ultra-microscope
6·11 Telescope

7. Speed of Light 80-89
7·1 Fizeau's Method
7·2 Foucault's Method
7·3 Speed of Light in a Transparent Liquid
7·4 Michelson's Method
7·5 The Kerr Cell Method

8. The Electromagnetic Theory of Light 90-104
8·1 Newton's Corpuscular Theory
8·2 Huygens' and Fresnel's Ether
8·3 Maxwell's Electromagnetic Waves
8·4 The Emission of Electromagnetic Waves

8·5 Energy Flux in an Electromagnetic Wave
8·6 Spatial Distribution of Dipole Radiation
8·7 Background of the Quautum Theory

9. Vibration and Waves **105-124**
9·1 Harmonic Oscillator
9·1A Use of Complex Amplitude
9·2 Damping in an Oscillator
9·2A Use of Complex Frequency
9·3 Fourier Series and Fourier Transform
9·3A Use of Exponential Representation
9·3B Fourier Transforms.
9·4 Wave Motion
9·5 Waves from Atoms and Molecules
9·6 Wave Groups and Wave Packets
9·7 Wave Velocity and Group Velocity
9·8 Direct Measurement of Wave Velocity
9.9. Necessity for Representation of Waves by Complex Quantities

10. Reflection and Refraction of Waves **125-135**
10·1 Huygens' Principle
10·2 Laws of Reflection from Wave Theory
10·3 Laws of Refraction from Wave Theory
10·4 Image Formation on Wave Theory
10·5 Simultaneous Reflection and Refraction at the Boundary of Two Media
10·6 Shock Waves : Cerenkov Radiation
10·7 Phase Change at Reflection
10·8 Ripple-tank Demonstration of Reflection and Refraction

11. Interference of Light [I] **136-152**
11·1 Theory of Interference
11·2 The Phasor Diagram Method
11·3 Interference in Space : Nodal and Anti-nodal Planes
11·4 Coherent Sources
11·5 Conditions for Interference of Light
11·6 Theory of Two Slit Interference
11·7 Young's Arrangement for Observing Interference
11·8 A Simple Equivalent to Young's Experiment

11·9 Fresnel's Biprism Arrangement
11·10 Llyod's Single Mirror Arrangement
11·11 Interference Fringes with White Light
11·12 Lateral Displacement of Fringes : Determination of Thickness of a Transparent Medium
11·13 Rayleigh Refractometer
11·14 Achromatic White Light Fringes

12. **Interference of Light [II]** 153-175
12·1 Multiple Reflection : Division of Amplitude
12·2 Interference in Thin Films
12·3 Colours of Thin Films
12·4 Newton's Rings
12·5 Haidinger's Fringes : Fringes in Parallel Films
12·6 Michelson's Interferometer
12·7 The Twyman and Green Interferometer
12·8 Multiple Beam Interferometry
12·9 Fabry and Perot Interferometer

13. **Diffraction of Light [I]** 176-192
13·1 Diffraction
13·2 Fresnel's Half-Period Zone Method
13·3 Fresnel's Diffraction at a Circular Aperture or Disc
13·4 Zone Plate
13·5 Explanation of Rectilinear Propagation
13·6 Diffraction at a Straight Edge
13·7 Cornu's Spiral

14. **Diffraction of Light [II]** 193-220
14·1 Fresnel and Fraunhofer Classes of Diffraction
14·2 Fraunhofer Diffraction at a Single Slit
14·3. Fraunhofer Dffraction as Fourier Transform of the Aperture
14·4 Fraunhofer Diffraction at a Circular Aperture
14·5 Emission by a Disc Source
14·6 Diffraction at Two Parallel Slits
14·7 Diffraction at *N* Parallel Slits
14·8 Plane Diffraction Grating
14·9 Plane Reflection Grating
14·10. Blazed Grating

14·11 Some Special Features of Grating Spectra
14·12 Concave Grating
14·13 Different Mountings of a Concave Grating

15. Resolving Power of Optical Instruments **221-242**

15·1 'Geometrical' and 'Spectral' Resolution
15·2 Distinction between 'Magnification' and 'Resolution'
15·3 Rayleigh Criterion for Limit of Resolution
15·4 Resolving Power of a Diffraction Grating
15·5 Resolving Power of a Prism
15·6 Some Devices for High Spectral Resolution
15·7 Resolving Limit of a Telescope
15·8 Michelson's Stellar Interferometer
15·9 Resolution in a Microscope
15·10 Resolution with Coherent Illumination : Abbe's Theory
15·11 Phase Contrast Microscope
15·12 Determination of Spectral Resolving Power
15·13 Determination of Resolving Power of a Telescope

16. Polarization of Light and Double Refraction **243-265**

16·1 Evidence that Waves of Light are Transverse
16·2 Polarization by Reflection : Brewster's Law
16·3 Polarization by Refraction through "Pile of Plates"
16·4 Double Refraction
16·5 Huygens' Explanation of Double Refraction
16·6 Electromagnetic Theory of Double Refraction
16·7 Experiment Determination of μ_0 and μ_E
16·8 The Polaroids
16·9 Nicol Prism
16·10 Double Image Prisms
16·11 Phase Retardation Plates
16·12 Elliptically and Circularly Polarized Light
16·13 Analysis of Polarization in a Given Beam of Light

17. Rolation of Plane of Polarization **266-278**

17·1 Optical Rotation
17·2 Laws of Rotation of Plane of Polarization
17·3 Origin of Optical Activity

17·4 Fresnel's Explanation of Optical Activity
17·5 Polarimeters
17·6 Optical Rotation by Magnetic and Electric Fields

18. Spectrographs and Various Kinds of Spectra 279-292
18·1 Littrow Spectrographs
18·2 Constant Deviation Spectrographs
18·3 Monochromators
18·4 Quartz Spectrograph
18·5 Spectrographs for Far Ultra-Violet
18·6 Spectrographs for the Infra-Red
18·7 Absorption Spectroscopy
18·8 Some Typical Spectra
18.9 A Note on Spectroscopic Measurements
18·10 Raman Spectroscopy
18·11 Luminescence Spectra

19. Origin of the Quantum Theory 293-315
19·1 Black-body Radiation and Planck's Theory
19·2 Photoelectric Effect and Einstein's Theory
19·3 Compton Effect
19·4 Series Relationship in Atomic Spectra
19·5 Energy Level Diagrams
19·6 The Franck and Hertz Experiments
19·7 The Specific Heats of Simple Solids
19·8 General Comments

20. The Shell Model of the Atom 316-338
20·1 Experiments on the Scattering of α-particles
20·2 Rutherford Model of the Atom
20·3 Bohr Model of the Hydrogen Atom
20·4 Excitation and Ionisation Potentials of Hydrogen
20·5 A Close Examination of Bohr's Theory
20·6 The Shell Model of the Atom
20·7 The Periodic Table of Elements
20·8 The Spectra of Alkali Atoms
20·9 The X-ray Spectra

21. Coherence and Lasers 339-362
21·1 Coherence : Purity of a Spectral Line
21·2 Stimulated or Induced Emission

21·3 Pumping and Population Inversion
21·4 Cavity Resonators and Shaping of a Beam
21·5 The Ruby Laser
21·6 Various Kinds of Lasers
21·7 High Monochromaticity of Laser Radiation
21·8 High Directionality of Laser Radiation
21·9 Uses of Lasers
21·10 Holography
21·11 Nonlinear Optics

22. Crystal Structure and Diffraction by Crystals 363-385
22·1 Space Lattices
22·2 Crystal Structure
22·3 Lattice Planes and Miller Indices
22·4 Diffraction of X-rays by Crystals
22·5 Diffraction of Material Particles
22·6 Applications of Electron and Neutron Diffraction

23. Elements of Quantum Mechanics 386-413
23·1 de Broglie Waves
23·2 Direct evidence for Wave Behaviour of Particles
23·3 Some Applications of de Broglie Relation
23·4 The Probability Interpretation of Phase Waves
23·5 Heisenberg's Uncertainty Relations
23·6 Some Applications of the Uncertainty Relations
23·7 The Schrodinger Wave Equation
23·8 Postulatory Basis of Quantum Mechanics
23.9 Some Applications of Schrodinger's Equation
23·10 The Hydrogen Atom

24. The special Theory of Relativity 414-433
24·1 Relativity and the Propagation of Light
24·2 The Michelson and Morley Experiment
24·3 The Lorentz Transformation
24·4 The Special Theory of Relativity
24·5 Length Contraction and Time Dilation
24·6 The Operational view of Time Dilation
24·7 Experimental verification of Time Dilation
24·8 The Relativistic Transformation of Velocities

24·9 Variation of Mass with Velocity
24·10 Relativistic Force and Kinetic Energy
24·11 Equivalence of Mass and Energy

25. Visual Photometry 434-448

25·1 Sensitivity of the Human eye
25·2 Luminous Intensity of a Source
25·3 Luminance and Brightness
25·4 Luminous Efficiency of a Source
25·5 Illuminance of a Surface
25·6 Typical Photometers
25·7 Brightness of Images
25·8 Viewing of Stars through a Telescope
25·9 The Human Eye

Answers to Numerical Problems 449-454

Index 455-463

1

Fermat's Principle and Ray Optics

Propagation of light in the early classes is treated with the concept of *rays*. But if you try a tight definition of 'ray' you get into trouble. We shall examine this aspect. Fermat gave a very great generalization, which not only gives all the laws of geometrical optics, but puts the 'ray' in the correct perspective : no ray can be considered in isolation. The law has a parallel in mechanics, which is outside our scope. But we will use some exercises to emphasize the crucial features.

1·1. A Close Examination of the Concept of Rays of Light. Among the simplest laws concerning the travel of light which we will examine are the following :

(*a*) *Law of mutual independence of rays of light passing through a given medium.*

(*b*) *Law of reversibility of the path of light.*

If several beams of light are passing through a given medium *simultaneously* in different directions, then the path of any beam is the same as it would be if all other beams were absent. That is the mutual independence referred to in law (*a*). The law of reversibility of paths means that if a ray *AB* follows a course *ABCDE*, then if we start with a ray *ED*, it must follow the *EDCBA*.

All such laws are good and valid *ordinarily*. The basic assumptions involved in all these laws are that light energy travels along a well defined path, which we call a ray. But when we take a closer look at these assumptions we face problems. In Fig. 1·1 consider a beam of light starting from *O*, spreading to width *PQ*, and then passing to the second medium on the right. In the figure we divide this beam into 5 parts, which are numbered 1·2...5. You may imagine the beam divided into 5000 parts in this way, or even

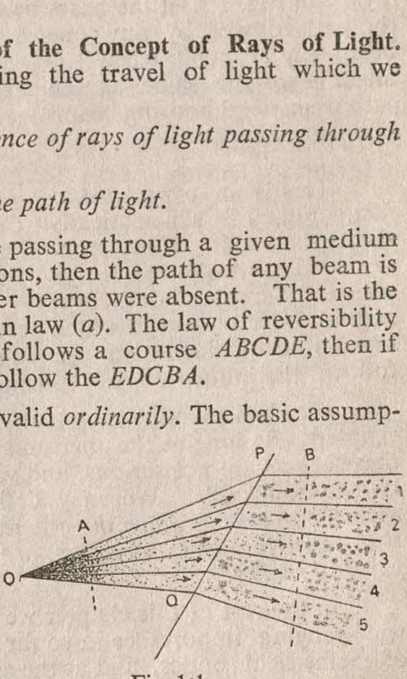


Fig. 1·1.
Spreading of a beam as it travels.

5×10^{20} parts if you like to get to your 'ray'. But you cannot ignore the fact that the beam *spreads as it travels.*

If this is the situation, what happens to the law of reversibility? At best it will have the sense shown in Fig. 1·2. The middle line (shown by dashes) in the left and right diagrams would match, but the spreads would not match. The spread speaks of the *time sequence* *A* to *E* for the first diagram and *E* to *A* for the second diagram.

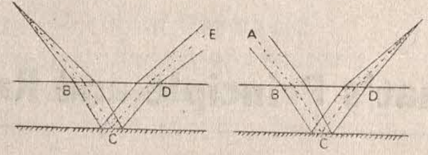


Fig. 1·2.

A close look at 'reversibility' of path of light.

Let us now have a look at the law of *mutual independence* of rays. In Fig. 1·3 let the beam be divided into, say, 100 parts. Then the path follwed by any one beam (say beam 50) when it is alone is *not* the same as the path when neighbouring beams are also present.

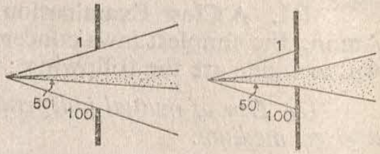


Fig. 1·3.

Course of 'ray number 50' in the presence of other neighbouring rays (left figure), and in the absence of other rays (right figure).

This is an *experimental fact*. Fig. 1·3 schematically shows the result. If ray number 50 alone is allowed to go ahead it spreads out, over and above the spread which occurred earlier. Thus *neighbouring rays* do have influence on the course which any given ray is to follow; the influence is of reducing any increase in the spread.

We shall examine such matters in quantitative details in later chapters. So long as the apertures of the beams are not too narrow, these effects are not serious, and we continue to treat the simple laws as applicable. We may define Geometrical 'Optics' as : *that branch of optics in which the geometrical laws of rectilinear propagation, reflection and refraction are good enough approximations to physical reality.*

In the rest of this chapter we shall discuss a principle of very wide-ranging importance in so far as it unifies in itself all the laws of geometrical optics, and at the same time also builds a bridge between geometrical and physical optics. This is Fermat's principle.

1·2. Fermat's Principle. As stated by Fermat in the year 1658 the principle is as follows :

A ray of light travelling from one point to another will follow a path such that, compared with neighbouring paths, the time required is the minimum.

This is called Fermat's Principle of Least Time. However, it is found that there are cases where the actual path is of maximum time among neighbouring paths. The principle is therefore modified to the following :

A ray of light travelling from one point to another will follow a path such that, compared with neighbouring paths, the time required in that path is a minimum, or maximum, or stationary.

In Fig. 1·4 P and Q are two points in different media. Let PAQ be a real path of light. Then time taken by light in reaching from P to Q along this path is given by

$$t=\frac{PA}{v_1}+\frac{AQ}{v_2} \quad \text{...(1·1)}$$

where v_1 and v_2 are the speeds of light in the two media.

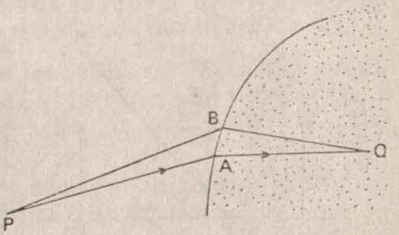


Fig. 1·4.
Showing a neighbouring path for application of Fermat's principle.

A *neighbouring path is* PBQ, and this may be expressed in terms of a parameter s, which could be either the lateral distance AB, or angle APB, or anything which shows a *lateral* shift of path. Fermat's principle now says that

$$\frac{dt}{ds}=0 \quad \text{...(1·2)}$$

Actually, in three-dimensional space a lateral shift of path can always be expressed in terms of two independent parameters s_1 and s_2. Therefore, Eq. (1·2) is equivalent to two independent conditions :

$$\left(\frac{\partial t}{\partial s_1}\right)_{s_2}=0 \quad \text{and} \quad \left(\frac{\partial t}{\partial s_2}\right)_{s_1}=0 \quad \text{...(1·3)}$$

In a medium of refractive index n the speed of light is given by $v=c/n$, where c is the speed in vacuum. Therefore, Eq. (1·1) may be expressed as

$$t=\frac{1}{c}(n_1PA+n_2.AQ) \quad \text{...(1·4)}$$

The quantity $n_1.PA + n_2.AQ$ is often called the *optical path*. If we represent it by p, then since c is constant, eq. (1·2) is equivalent to

$$\frac{dp}{ds}=0 \quad \text{...(1·5)}$$

This states Fermat's principle in terms of the *optical path length* p, instead of *time* t. The *optical path* in general is $\Sigma n_i l_i$ where n_i is the refractive index of the medium for the i^{th} part of the path of length l_i.

Example 1·1. Let A and B be two point in a plane as shown in the Fig. 1·5. P is a variable point defined by $CP=s$, and $AP+PB$ is a variable path length p dependent on s. Take $s=0, 10, 20,\ldots 70$ units, compute p each time, and draw a graph for p versus s. Find if there is a path which is maximum or minimum amongst neighbouring paths.

Solution. From geometry of the figure

$$p=[30^2+s^2]^{1/2}+[50^2+(70-s)^2]^{1/2}$$

Straight computation gives the following :

$s=$	0	10	20	30	40	50	60	70
$p=$	116·1	109·7	106·8	106·4	108·3	112·1	118·1	126·2

The graph between p and s may now be drawn. We get a *minimum* p at $s=26$. ⊙

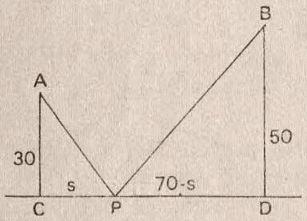


Fig. 1·5.
The geometry for example 1·1.

Example 1·2. Reverting to the situation of example 1·1, let 70 messengers start simultaneously from A to reach B through paths with $s=1, 2,\ldots 70$. If the speeds of all are equal, being 10 speed units, deduce from the plot of t versus s how many of the messengers reach within 0·1 time unit of the arrival of the messenger reaching earliest.

Solution. Converting p/speed=time t, the data from solution of Example 1·1 become the following :

$s=$	0	10	20	30	40	50	60	70
$t=$	11·61	10·97	10·68	10·64	10·83	11·21	11·81	12·62

The graph (not shown) gives minimum time=10·62 units (for messenger number 26). Within 0·1 unit of time 18 other messengers reach B (messenger numbers 17 to 35). [Note that in the next 0·1 unit time interval only 5 or 6 messengers would reach.] ⊙

Example 1·3. In Example 1.2, suppose the messengers have to pick some load from positions s so that speed over distance PB is reduced to half, *i.e.*, 5 speed units. Which messenger now will have the shortest timing ?

Solution. Now $t=\frac{1}{10}[{}^2+s^2]^{1/2}+\frac{1}{5}[50^2+(70-s)^2]^{1/2}$

Proceeding as for Example 1·1 shows that the messenger with least time among neighbours is now number 47 (from graph) with timing 16·59 units. ⊙

The purpose of the sequence of examples given above is to show the meaning of Fermat's principle in depth. The 'messengers' of examples 1·2 and 1·3 are the light signals starting out from A. The real signals reaching B are those which have crowding of neighbouring signals in time. This condition is satisfied at $dt/ds=0$. One may check also that in Example 1·1 for $s=26$ the angles which AP and PB make with the normal to CD are equal, and in Example 1·3 for $s=47$ the ratio of sines of these angles is 2, equal to the ratio of the speeds.

Fermat's principle in effect tells us that no ray is to be considered in isolation. It specifically asks us to extend our examination to the neighbourhood of the considered path (or ray), to check if timings for those paths would *crowd together* around that for the path under considertion. If we can establish the answer as 'yes' then alone the path considered becomes a real path.

In the wave theory of light, as you will see later, the equivalent question is whether contributions from the neighbouring parts of the wavefront reach *B* in agreement of *phase*.

1·3. Laws of Reflection from Fermat's Principle. In Fig. 1·6 let *JKLM* be a plane reflecting surface and *A* and *B* two points outside this plane. Rays from *A* falling over various points on *JKLM* go in various different directions, and we wish to enquire which one of these rays would in practice pass through B.

Draw normals *AO* and *BO′* on the reflector. Take the origin of coordinates at *O* and let *OO′* and *OA* be the *x* and *y* axes. The *z* axis is then perpendicular to *OO′* and lies in the plane of the reflector.

If $OA=a$, $O'B=b$ and $OO'=c$, then our choice of axes gives the coordinates of *A* as $(0, a, 0)$ and of *B* as $(c, b, 0)$.

Now imagine any general point *P* on the reflector. The *y*-coordinate for it will be 0, and *x* and *z* coordinates will be variable. So *P* has general co-ordinates $(x, 0, z)$.

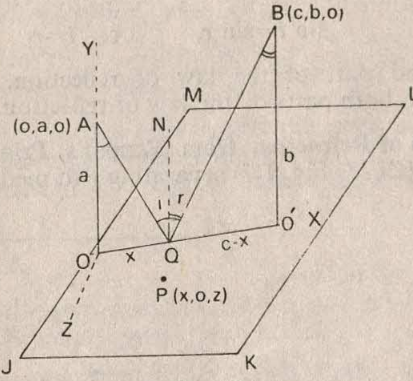


Fig. 1·6.

Various paths from A to B via surface JKLM.

APB is now a most geenral *conceivable* path from *A* to *B* via the reflector. If we set $AP+PB=p$, and speed of light in the medium is *v*, we get

$$p=[(x-0)^2+(0-a)^2+(z-0)^2]^{1/2}+[(x-c)^2+(0-b)^2+(z-0)^2]^{1/2}$$

$$t=\frac{p}{v}=\frac{1}{v}(x^2+a^2+z^2)^{1/2}+\frac{1}{v}[(x-c)^2+b^2+z^2]^{1/2} \qquad \ldots(1{\cdot}6)$$

where t is the time taken over the path.

We now apply Fermat's principle to get the real path. The path can be varied by varying z and x, which are the free coordinates for P. Hence the conditions for real path are

$$\left(\frac{\partial t}{\partial z}\right)_x=0 \quad \text{and} \quad \left(\frac{\partial t}{\partial x}\right)_z=0 \qquad ...(1\cdot7)$$

Deducing $(\partial t/\partial z)_x$ from Eq. (1·6) and setting this equal to zero gives at once

$$z=0 \qquad ...(1\cdot8)$$

This means that P must lie in the plane $AOO'B$, which is normal to the reflector. If Q is such a position for P, then incident ray AQ, the normal QN and the reflected ray QB all lie in the same plane. Thus, Eq. (1·8) gives the first part in the law of reflection.

Now, substituting $z=0$ in Eq. (1·6), deducing $(\partial t/\partial x)z$, and setting this equal to zero (see Eq. 1·7) gives

$$\frac{c-x}{[(x-c)^2+b^2]^{1/2}}=\frac{x}{(x^2+a^2)^{1/2}} \qquad ...(1\cdot9)$$

If we look at Fig. 1·6 the two sides of Eq. (1·9) are simply equal to sin r and sin i, where i and r are the angles of incidence and of reflection respectively. Thus

$$\sin i=\sin \mathbf{r}, \qquad \text{or} \quad i=r \qquad ...(1\cdot10)$$

This is the second part of the law of reflection. Thus Fermat's principle leads to both parts of the law of reflection.

1·4. Laws of Refraction from Fermat's Principle. Consider a plane surface $JKLM$ (Fig. 1·7) separating two media. In the upper

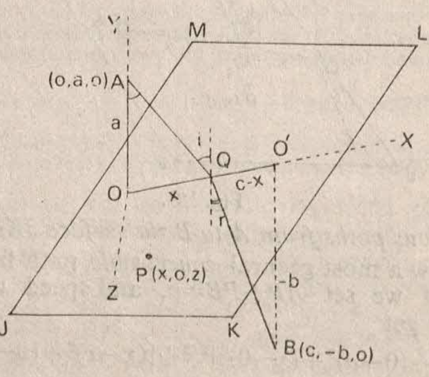


Fig. 1·7.

The most general path from A to B is APB.

medium the speed of light is v_1 and in the lower one v_2. If A and B are points lying in the two different media, what is the *actual* path of a ray from A to B?

Proceeding exactly as for Eq. (1·6), the travel time t for a path $AP+PB$ via a general point (x, 0, z) is now given by

$$t=\frac{AP}{v_1}+\frac{PB}{v_2}=\frac{1}{v_1}(x^2+a^2+z^2)^{1/2}$$
$$+\frac{1}{v_2}[(x-c)^2+b^2+z^2]^{1/2} \qquad \ldots(1{\cdot}11)$$

Now we apply Fermat's principle that varying the position P should not vary t in the neighbourhood of the actual path, *i.e.*,

$$\left(\frac{\partial t}{\partial z}\right)_x=0 \text{ and } \left(\frac{\partial t}{\partial x}\right)_z=0$$

The first of these, applied to Equation (1·11) gives us $z=0$, which is the first part of the law of refraction. Then the second condition gives us

$$\frac{x}{(x^2+a^2)^{1/2}}\div\frac{(c-x)}{[(c-x)^2+b^2]^{1/2}}=\frac{v_1}{v_2} \qquad \ldots(1{\cdot}12)$$

A look at Fig. 1·7 at once shows that Eq. (1·12) actually states the Snell's law of refraction, *viz*.

$$\frac{\sin i}{\sin r}=\frac{v_1}{v_2}=\text{constant} \qquad \ldots(1{\cdot}13)$$

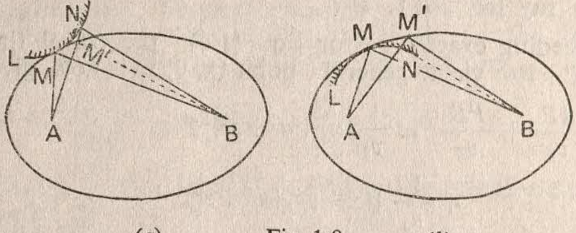
This equation further tells us that the constant, which we call the refractive index and designate n_{12}, is equal to the ratio of the speed of light in the first medium to that in the second one, *i.e.*,

$$n_{12}=\frac{v_1}{v_2} \qquad \ldots(1{\cdot}14)$$

Note that this information (1·14) is not contained in Snell's Law ; Fermat's principle gives it as a bonus.

1·5. Extremum Path ; not Least Path. We have already stated that Fermat's principle is not the principle of least time (or least optical path) only, but one of *least or greatest or stationary* time among the times for neighbouring paths. We shall now show expressly the conditions under which a path may be least or greatest among neighbouring paths. We shall take up the case of reflection, leaving that of refraction for the reader to try out.

In Fig. 1·8(*a*) consider A and B as two points and LMN a curved mirror. Let AMB be the *actual* path of a ray from A reaching B after reflection at this mirror. Is this path a maximum or a minimum among neighbouring paths ? This can be settled as follows:



(*a*) Fig. 1·8. (*b*)

Showing Fermat's law of extremum path for case of reflection,

Draw an ellipse passing through M and having A and B as its focii. Rotate this ellipse about AB as axis. We get an ellipsoid. By property of an ellipse, $AM+MB$=constant for all positions of M on this ellipse. For example,

$$AM+MB=AM'+M'B.$$

Now consider the actual mirror LMN, which in Fig. 1·8(*a*) falls *outside* the ellipsoid just drawn. A path neighbouring to AMB is ANB.

We note that difference of path is

$$\begin{aligned}ANB-AMB&=(AM'+M'N+NB)-(AM+MB)\\&=(AM'+M'N+NB)-(AM'+M'B)\\&=M'N+NB-M'B.\end{aligned}$$

This is always positive, since $M'N$, NB are two sides of a triangle of which $M'B$ is the third side. Thus any neighbouring path (ANB) is *longer* than the actual path (AMB) for this case.

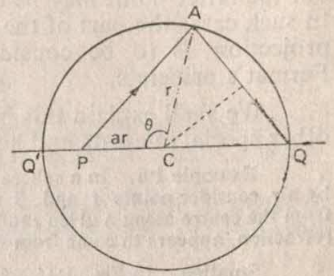
Now consider Fig. 18(*b*), in which the given mirror surface falls *inside* the ellipsoid passing through M and having A, B as focii. Here

$$\begin{aligned}ANB-AMB&=(AM'-M'N+NB)-(AM+MB)\\&=(AM'-M'N+NB)-(AM'+M'B)\\&=-(M'N+M'B-NB)\end{aligned}$$

This is always negative. Hence in this case, any neighbouring path (ANB) is *shorter* than the actual path (AMB).

Thus, the actual path is minimum among the neighbours if the reflecting surface touches the locus surface (the ellipsoid) from outside, and it is maximum among the neighbours if the reflecting surface touches the locus surface from inside, and it is stationary among the neighbours if the reflecting surface coincides with the locus surface $AM+MB=$constant.

Example 1·4. In Fig. 1·9 P is a point source distant 0·6 r from the centre C of a spherical reflecting surface of radius r. From Fermat's Law show that if ray PA is reflected to Q, the angle θ is given by $\cos^2\theta/2=2/3$.



Fig, 1·9.
The geometry for Example 1·4.

Solution. In general, let $PC=a.r$. Then $p=PA+AQ$

$$=r(1+a^2-2a\cos\theta)^{1/2}+2r\cos(\theta/2)$$

Applying Fermat's law $\partial p/\partial\theta=0$ leads to two solutions

$$\sin\theta/2=0:$$

$$2a\cos\theta/2=(1+a^2-2a\cos\theta)^{1/2}$$

The first gives one possible incident ray as PQ', which after reflection passes through Q. The second one leads to

$$\cos^2(\theta/2)=\frac{1+a}{4a} \qquad \ldots(1\cdot15)$$

In the given case $\qquad a=0\cdot6.$

$$\therefore\ \cos^2\theta/2=\frac{1\cdot6}{2\cdot4}=2/3$$

[Note : If $a<1/3$, the only ray from P which can reach Q is PQ'.] ⊙

Example 1·5. In Fig. 1·10, C is the centre of a refracting sphere of radius r, refractive index $n=1\cdot50$. A parallel beam is incident in the direction OC. Find out that ray which after refraction passes through a point Q distant $CQ=1\cdot2\ r$.

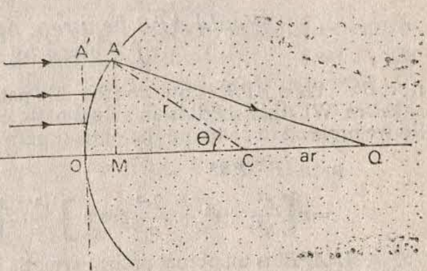


Fig. 1·10.
The geometry for Example 1·7.

Solution. For all incident rays the path up to the dotted line OA' is common. So the variable path is $p=A'A+n.\ AQ$. The parameter convenient to take is $\angle OCA=\theta$.

$$p=A'A+nAQ=r(1-\cos\theta)+nr\,(1+a^2+2a\cos\theta)^{1/2}$$

Applying Fermat's law $\partial p/\partial\theta=0$ leads to two solutions :

Either $\qquad \sin\theta=0$

or $\qquad (1+a^2+2a\cos\theta)^{1/2}=na$

This first solution gives the ray incident along OC, which passes through all positions of Q on line OC. The second gives

$$\cos\theta=\frac{(n^2-1)a^2-1}{2a} \qquad \ldots(1\cdot16)$$

In the present case $n=1\cdot50$, $a=1\cdot2$

$$\therefore\quad \cos\theta=\frac{1}{3}\ \text{(check it)}$$ ⊙

1·6. Case of a Virtual Point on the Path. A case may arise where a ray starting from a given point A after suffering a series of reflections, refractions, etc., may not pass through another point B, but the latter point may be on the final path produced backwards. In such cases the part of the path which is obtained by backward projection is to be considered with negative sign in applying Fermat's principle.

We shall explain this by using a concrete example, which has other special interests and has important practical applications too.

Example 1·6. In a sphere of radius r and refractive index n surrounded by air, consider points A and B situated at distance r/n and rb respectively from the centre along a given radial line. Deduce that ray from A which, after refraction, appears to come from B.

Solution. In Fig. 1·11, $OA=r/n$ and $OB=rb$. Now consider a point P on the sphere specified by parameter θ. We join AP, and join BP, and produce, it to B'. Then APB' is one conceivable path between A and B, with part PB *negative*. Whether APB' is an actual path, only Fermat's principle will tell us. So we proceed to compute the optical path $APB=p$.

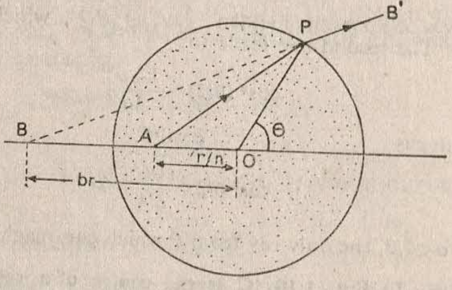


Fig. 1·11.

Would APB be a real path for any position of B given by OB=br ?

$$p=nAP-PB^*$$

$$=n\left[\frac{r^2}{n^2}+r^2+\frac{2r^2}{n}\cos\theta\right]^{1/2}-\left[r^2b^2+r^2+2r^2b\cos\theta\right]^{1/2} \qquad \ldots(1\cdot17)$$

If APB' is an actual path, Fermat's law demands $dp/d\theta=0$, since θ is the variable parameter here. That leads to one condition $\sin\theta=0$, which is trivial. The other condition, on simplification, comes to

$$\frac{1}{n^2}+\frac{2}{n}\cos\theta=\frac{1}{b^2}+\frac{2}{b}\cos\theta \qquad \ldots(1\cdot18)$$

Now, for a general value of b this condition is not satisfied for any value of θ. This is a new situation. No ray from A, whatever its point of incidence P at the sphere, can pass through the point specified by $OB=rb$.

If, however, the point B is specified by $OB=nr$ (*i.e.*, we have $b=n$), then Eq. (1·18) is satisfied. But again there is a new situation. Eq. (1·18) is now satisfied for *all values* of θ. Hence, not only a particular ray but *all the rays* from *A will after refraction have B on their backward projection.*

---

*The virtual path is taken as traversed in the second medium (here air), although in the figure a part of the dotted line BP appears in the first medium.

This result is used in making aplanatic lenses and in oil immersion objectives of microscopes. (Chapter 5) ⊙

PROBLEMS

1·1. Laws of geometrical optics are said to be applicable when width of the beam is large. 'Large' compared to what ? Discuss. (See Fig. 1·3 right).

1·2. State Fermat's principle and explain the significance of the neighbouring paths in this law. (see end part of §1·2, and discussion with Fig. 1·5).

1·3. Deduce the conditions of geometary which decide whether the actual path of a ray is a maximum or minimum path among neighbouring paths (§ 1·5)

1·4. Using the theorem in geometry that sum of two sides of a triangle is greater than the third side, give a formal proof of rectilinear propagation of light, based on Fermat's principle.

1·5. Establish the condition of total internal reflection using Fermat's principle directly. (Prove absence of refracted ray if $v_2>v_1$ and sin $i>v_1/v_2$ in Fig. 1·7.

1·6. Consider a situation in the atmosphere when the refractive index increases quite rapidly as we go up. Discuss in terms of Fermat's principle the *two* possible paths of rays from a low point at one place to a high point at large horizontal distance. [Hint : One path is of refraction, other for total reflection.]

1·7. In the diagram below, AO is a spherical surface of radius r separating media of refractive indices n_1 and n_2. P_1 and P_2 are points on the principal axis at distances a_1 and a_2 from pole O. Show that (i) the path P_1OP_2 is always an actual path of rays, and that among neighbouring paths this path P_1OP_2 is (ii) a maximum for $r>r^\circ$ and (iii) minimum for $r<r_0$, where r_0 is given by

$$\frac{1}{r^\circ}=\frac{1}{n_2-n_1}\left(\frac{n_1}{a_1}+\frac{n_2}{a_2}\right)$$

[*Hint* : Deduce expression for the path in terms of parameter θ, and then examine the first two derivatives.]

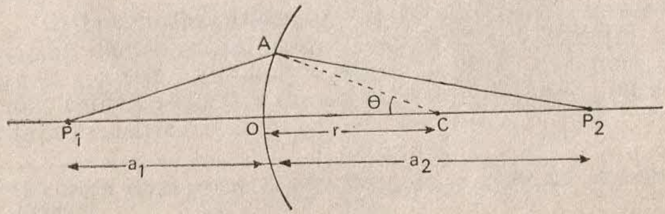


Fig. 1·12.

1·8. Show that eq. (1·16) leads to $a \to (n-1)^{-1}$ and $(n^2-1)^{-1/2}$ as $\theta \to 0°$ and $90°$ respectively. Interpret the first result.

1·9. Parallel to Example 1·5, discuss the case when the refracting surface AO is concave.

1·10. Consider points P and Q at opposite ends of a diameter of a spherical reflector. Deduce the possible actual paths for rays of light reaching Q from P, stating which path corresponds to a maximum among neighbours and which to a minimum.

1·11. In Example 1·4 (Fig. 1·9) show that the path defined by $\cos^2 \theta/2 = 2/3$ is a minimum among neighbouring paths.

1·12. Using Fermat's law deduce the focal length of a convex mirror. [Hint ; Use Fig. 1·10 with path AQ taken negative and a as variable. Result $|f| = r(1 - \frac{1}{2} \sec \theta)$].

2

Cardinal Points of an Optical System

In this chapter we shall deal with another great generalisation in the laws of optics. This relates to image formation by any complex optical system, and the generalization was given by Gauss. If we have a succession of reflecting or refracting surfaces 1, 2, 3,...*j*. each one of which forms an image under well defined rules, then applying *j* relations in succession to find the position and size of the final image would be tedious and complex. Is it possible to represent the action of all these *j* surfaces taken together with the help of a very few parameters, which in a single step give us the final image of the object ? Gauss answered this question in the affirmative. That was a great step indeed—great saving of labour.

But there is something more in Gauss's work. It leads to certain very general results which do not depend on the system chosen. These are something like the laws of thermodynamics, which are valid for all systems.

We shall try to give only a rapid scan of the great generalization, dropping out some finer details of rigour. From time to time we shall take examples from real simple systems, but the beauty of the theory is in its universal validity.

2·1. Convention of Signs. In dealing with the theory of image formation, a definite convention regarding signs is needed. The convention we adopt in this book is as follows :

(i) Axes. The principal axis is chosen as *x*-axis. The y-axis is perpendicular to this in the plane of the drawing. (Due to axial symmetry, *z*-axis need not be defined unless we consider aberrations.)

(ii) Longitudinal Distances. *Object distances* measured (from a suitably chosen origin) in the direction of the incident rays are taken as positive. Similarly, *image distances* measured (from a suitably chosen origin*) in the direction of the reflected/refracted ray are taken positive.

---

*This origin will in general be different from the origin for measuring object distances.

(iii) Radius of Curvature of a surface is taken positive if it is convex towards the incident rays.

(iv) Lateral Distances. These are measured always from the x-axis, and taken positive upward and negative downward.

(v) Angles with the axes. Slope angles of rays with the axes are measured between the positive directions of the axis and the ray, and are considered positive if the *axis* needs an anticlockwise turn to bring it in line with the *ray*.

(vi) Angles of incidence, reflection and refraction are taken positie if the *normal* at the point under consideration is to be turned anticlockwise to coincide with the *positive* direction of the *ray* under consideration. The normal which makes an acute angle with the positive direction of the ray is considered for the purpose.

Fig. 2·1 explains the convention of signs. Let H and H' be the origins for object and image spaces respectively. PP' is the object, and QQ' the image.

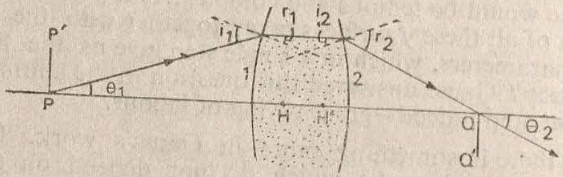


Fig. 2·1.

Explaining the convention of signs.

Convention (ii) makes H to P negative (opposite to ray direction) and H' to Q positive.

Convention (iii) makes R_1 positive and R_2 negative.

Convention (iv) makes PP' positive and QQ' negative.

Convention (v) makes θ_1 positive and θ_2 negative.

Convention (vi) makes i_1 and r_1 positive, and i_2 and r_2 negative.

2·2 Gauss's General Theory of Image Formation in a Coaxial Symmetric System. Gauss's theory deals with ideal image formation, by which we mean cases where there is a point-for-point, line-for-line and plane-for-plane relationship between an object and its image. With the various refracting or reflecting surfaces coaxial the common axis is taken as x-axis. Due to symmetry about this axis, it is enough to deal with any one plane through the x-axis, and transverse distances give y.

With arbitrary choice of origins for the object and image space, let (x,y) be the coordinates of an object point and (x', y') those for its image. Then the conditions of ideal image formation are satisfied if

$$x'=\frac{a_1x+b_1y+d_1}{ax+by+d}$$
$$y'=\frac{a_2x+b_2y+d_2}{ax+by+d} \qquad \text{...(2·1)}$$

We state this without a rigorous mathematical proof. We can see that out of the nine constants involved, five are redundant. If y changes sign, x' should remain unaffected. This makes $b=b_1=0$.

Again, if y changes sign, y' should only change in sign, not in magnitude. This makes $a_2=d_2=0$. Finally, dividing all coefficients by a, and expressing the new values by the corresponding capital letters, we get

$$\boxed{x'=\frac{A_1x+D_1}{x+D}, \quad y'=\frac{B_2y}{x+D}} \qquad \text{...(2·2)}$$

Thus we reach the very important conclusion that *in any co-axial symmetrical system forming ideal images the co-ordinates x', y' of an image point corresponding to the object point x, y are uniquely determined by four and only four independent constants*, however complex the system may be.

In actual practice, two of these constants are used to specify the origins for object space and image space on the X-axis for measurement of distances. Then only two other constants specify the object to image conjugate relations uniquely. There are several alternative ways of choosing the origins, and we shall discuss two of them in the following sections.

2·3. Conjugate Relationships. By definition, *the first focal point of the system is at that value of x which makes $x'=\infty$.* Hence from Eq. (2·2), we get

$$\textbf{first focus} \quad \text{at } x=-D \qquad \text{...(2·3)}$$

Similarly, putting $x=\infty$ in Eq. (2·2), we get

$$\textbf{second focus} \text{ at } x'=A_1 \qquad \text{...(2·4)}$$

We now transfer the origins for measurement of x and x' to these corresponding *focal points*. Using suffix o for the new coordinates, the transformations are

$$\left.\begin{matrix} x_o=x+D & y_o=y \\ x_o'=x'-A_1 & y_o'=y' \end{matrix}\right\} \qquad \text{...(2·5)}$$

Substituting in Eq. (2·2) and writing D_1-A_1D as a new constant A we get

$$\left.\begin{matrix} x_o\,x_o'=A \\ \dfrac{y_o'}{y_o}=\dfrac{B_2}{x_o} \end{matrix}\right\} \qquad \text{...(2·6)}$$

Now only two constants A and B_2 specify the relations between conjugate object points and image points of the given system. The orgins are now not arbitrary, but are fixed at the two focal points.

Principal Planes ; Focal Lengths. Principal (or unit) planes are that pair of conjugate planes for which the lateral magnification $m=y_o'/y_o$ is unity. The points where these planes meet the axis are called the Principal (or unit) points. Eqs. (2·6), with $y_o'/y_o=1$, tell us that the principal points (H and H') are located at

$$\left.\begin{array}{l} x_o \text{ (for } m=1)=B_2 \text{ (distancc } F \text{ to } H) \\ x_o' \text{ (for } m=1)=A/B_2 \text{ (distance } F' \text{ to } H') \end{array}\right\} \quad \cdots(2{\cdot}7)$$

The distance *from the first principal point* H to the first focus F (*i.e.*, $-B_2$) is called the focal length of object space f. Similarly, the distance from the second principal point H' to the second focus F' (*i.e.*, $-A/B_2$) is called the focal length of image space f'.

Substitutions $f=-B_2$ and $f'=-A/B_2$ in Eq. (2·6) lead to

$$\boxed{x_o x_o'=ff'}^{*} \quad \cdots(2{\cdot}8a)$$

$$\boxed{\frac{y_o'}{y_o}=-\frac{f}{x_o}=-\frac{x_o'}{f'}} \quad \cdots(2{\cdot}8b)$$

The notable fact is that we did not use any particular system to deduce the results. We only used the assumption that ideal image formation is possible. Then the system, however complex, must have a pair of focal points (F,F') and a pair of focal lengths (f,f'). With x_0 and x_0' measured from F and F' respectively, Eqs. (2·8) give the conjugate relationship and lateral magnification.

Example 2·1. In a co-axial optical system using an arbitrary origin for expressing distances, the conjugate relations were found to be

$$x'=\frac{20x+288}{x+15},\ y'=\frac{3y}{x+15}$$

Deduce the positions of F, F', H, H' and values of f, f'.

Solution. Setting $x'=\infty$ gives position of F $x=-15$
Setting $x=\infty$ gives position of F' $x'=20$
Setting $y'=y$ gives position of H $x=12$
Setting $x=-12$ gives position of H' $x'=16$
f = distance H to $F=x(F)-x(H)=-3$
f' = distance H' to $F'=x'(F')-x'(H')=4$

[Note that x and x' values are from arbitrary origins. But f and f' values, being differences, are unique for the system.]

⊙

Example 2·2. In a co-axial optical system the focal points are located by an experiment. Then for $x_0=-6$ cm we find $x_0'=12$ cm and $y_0'/y=-3/2$. Deduce f and f'. Also predict x_0' and magnification m for $x=-10$ cm.

Solution. Use of Equation (2·8b) gives

$$f=-9 \text{ cm.} \qquad f'=+8 \text{ cm.}$$

---

* This is called Newton's formula. For most practical cases $f'=-f$. Hence, dropping subscript 0, we get the form $xx'=-f^2$.

For the case $x_0=-10$ cm, straight substitution gives

$$x_0'=+7\cdot2\text{ cm.}\quad m=-0\cdot9. \qquad \odot$$

Relationships with Principal Points as Origins. Let us now decide to measure object and image distances from the respective principal points. If these distances are called u and v, the transformations are

$$\left.\begin{aligned} u&=x_o+f \qquad & y&=y_o \\ v&=x_o'+f' \qquad & y'&=y_o' \end{aligned}\right\} \qquad \ldots(2\cdot9)$$

Equation (2·8) now give*

$$\boxed{\frac{f}{u}+\frac{f'}{v}=1} \qquad (\ldots2\cdot10a)$$

$$\boxed{m=\frac{f}{-u}=\frac{f'-v}{f'}} \qquad \ldots(2\cdot10b)$$

Example 2·3. In a co-axial system of lenses, the first and secone focal points fall at 10·0 and 22·7 cm. and the principal points at 17·7 and 15·0 cm. respectively, all measured from a common origin. Deduce (i) thd focal lengths of object space and image space, (ii) the position of the imaga of an object placed at position −2·7 cm.

Solution. The locations are

$F(10\cdot0)$, $H(17\cdot7)$, $F'(22\cdot7)$ and H' $(15\cdot0$ cm$)$.
f=distance from H to $F=10\cdot0-17\cdot7=-7\cdot7$ cm.
f'=distance from H' to $F'=22\cdot7-15\cdot0=7\cdot7$ cm.
x=distance from F to $-2\cdot7=-2\cdot7-10\cdot0=-12\cdot7$ cm.
$x'=ff'/x=(-7\cdot7)\times7\cdot7/(-12\cdot7)=+4\cdot7$ cm.

Since this is the distance of image from the second focus, the position of image is $22\cdot7+4\cdot7=27\cdot4$ cm.

2·4. Cardinal Points of an Optical System.

In any system of lenses or mirrors or both, which forms ideal images of objects, the focal points (F,F') and the principal points (H,H') taken together are called the *cardinal points* of the system.** If these are given, then the image point Q corresponding to any object point P can be determined either from eqs. (2·8) or (2·10), or from geometrical drawing.

Fig. 2·2 explains the latter procedure. From P draw a ray PF passing through F, and project it till it meets the first principal plane at C. Now we note that the conjugate of point C must lie in the second principal plane such that $H'C'=HC$. This gives us the

---

*Note the symmetry in eq. (2·10*a*). With $f=-f'$, it gives the form $\frac{1}{v}-\frac{1}{u}=\frac{1}{f'}$. But that is not the universally applicable form.

**The nodal points are two additional cardinal points often useful. The nodal points are defined as that pair of conjugate points in object and image space which correspond to + 1 *angular* magnification. However, we shall not discuss them further.

point C'. We also know that since ray PFC passes through the first focus F, its conjugate must be parallel to the principal axis. So from

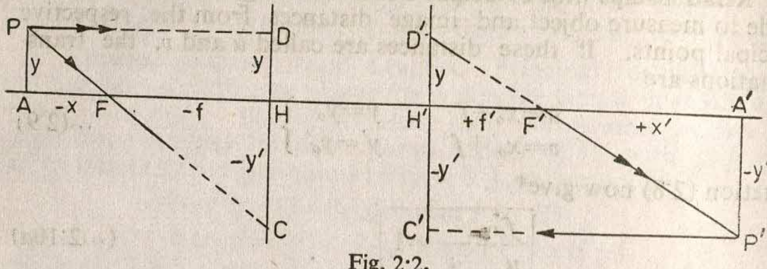


Fig. 2·2.

Geometrical Deduction of Image

C' we draw $C'P'$ parallel to the principal axis.

Next, from P we draw PD parallel to the principal axis, meeting the first principal plane at D. The conjugate of D is D', obtained by the rule that it must be in the second principal plane, with $H'D'=HD$. Further, since ray PD is parallel to the principle axis, its conjugate must pass through F'. So from D' we draw a ray passing through F' and produce it till it meets the first ray at P'.

Corresponding to the two incident rays PC and PD, the emergent rays meet at P'. Since the system is assumed to give ideal images, all rays starting from P must on emergence pass through P'. Thus P' is the conjugate of object point P. It may be noted that this tracing required all the *four* points F,F', H,H', ; less will not do and more information is not needed.*

2·5. Actual Systems. The general treatment given so far does not tell us where F,F' lie in any actual system, and whether focal lengths f and f' bear any relation to one another. To determine these inter-relationships we shall require the application of the *actual* laws of refraction (or reflection) in some generalized form.

However, before we do that it will be useful to examine some familiar simple cases like those of single refraction and a thin lens to discover if they indicate some simple relationships.

(a) Refraction at a Spherical Surface. In Fig. 2·3, OA represents the spherical surface of separation of two media of refractive indices n_1 and n_2. C is the centre of the surface, P is an object point, PC meets the spherical surface at O, which gives the pole. For a ray PA, the normal is CAN, and the refracted ray is AQ.

---

* We may use Fig. 2·2 to deduce Eqs. (2·8) and (2·10). This is left as an exerise for the reader.

Using Snell's law $n_1 \sin i_1 = n_2 \sin r_2$ (with first order approximation $n_1 i_1 = n_2 r_2$) and the geometry of Fig. 2·3 one deduces*

$$n_1 \angle APO - n_2 \angle AQO = (n_1 - n_2) \angle ACO$$

or

$$\frac{n_2}{v} - \frac{n_1}{u} = \frac{n_2 - n_1}{R} \qquad \ldots(2{\cdot}11)$$

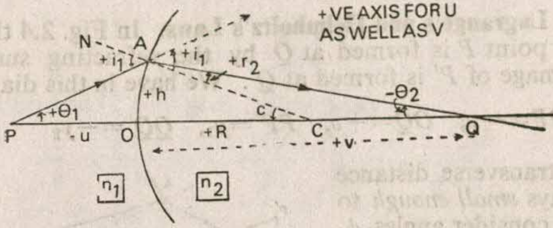


Fig. 2.3

One example of $f'/f = -n_2/n_1$

Now, if we put $v = \infty$, the corresponding u is, by definition, the *focal length of object space f.* If we put $u = \infty$, the corresponding v is by definition the *focal length of image space f′*. We thus get

$$f = -\frac{n_1 R}{n_2 - n_1}; \quad f' = +\frac{n_2 R}{n_2 - n_1} \qquad \ldots(2{\cdot}12)$$

We note that f and f' are related by

$$\boxed{\frac{f'}{f} = -\frac{n_2}{n_1}} \qquad \ldots(2{\cdot}13)$$

Relations (2·12) give f and f' for the particular system in terms of n values and R. But Eq. (2·13) may have a more *general* validity.**

(b) Thin Lens in Any Medium. Consider a thin lens of radii of curvature R_1, R_2 and refractive index n. For generality, let the initial and final media have refractive indices n_1 and n_2 respectively. Then applying Eq. (2·11) in two steps, one gets

$$\frac{n_2}{v} - \frac{n_1}{u} = n\left(\frac{1}{R_1} - \frac{1}{R_2}\right) + \left(\frac{n_2}{R_2} - \frac{n_1}{R_1}\right) = \frac{1}{G} \text{(say)}, \ldots(2{\cdot}14)$$

where G is a constant of the system having the dimensions of length. On applying the definition $u \equiv f$ when $v = \infty$, *and* $v \equiv f'$ when $u = \infty$, we get

---

*Note that in the first step only the geometry comes; in the second step the algebraic signs for θ_1, θ_2, v and u are used. That practice is to be used all along.

**Note that the essential element introduced in the deduction is Snell's law, all the rest is pure geometry.

$$f=-n_1G \text{ and } f'=+n_2G \qquad \cdots(2{\cdot}15)$$

so that again

$$\frac{f'}{f}=-\frac{n_2}{n_1} \qquad \cdots(2{\cdot}16)$$

A general proof of this will be given in § 2·7.

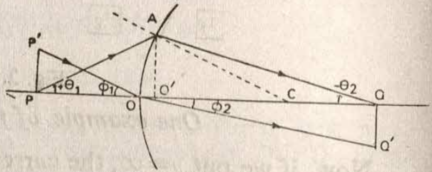
2·6. Lagrange's and Helmholtz's Laws. In Fig. 2.4 the image of an axial point P is formed at Q by the refracting surface OA. Also, the image of P' is formed at Q'. We have in this diagram

$$OP=-u, \quad OQ=+v, \quad PP'=y_1, \quad QQ'=-y_2$$

The transverse distance PP' is always *small enough* to allow us to consider angles ϕ_1 and ϕ_2 small, so that, using Snell's law, weget

$$\frac{n_1}{n_2}=\frac{\sin\phi_2}{\sin\phi_1}$$

$$\approx\frac{(-y_2)}{v}\div\frac{y_1}{(-u)}=\frac{y_2}{y_1}\cdot\frac{u}{v} \qquad \cdots(2{\cdot}17)$$



Eig, 2.4.
For deduction of Lagrange's law of magnification

The transverse distance OA, which relates to the aperture of the system, is not always so small that we may put $\tan\theta_1=\theta_1$. etc. However, if O' is the base of perpendicular from A on the axis, OO' is still small enough to be neglected (as compared to u and v), so that we may put $O'P=-u$ and $O'Q=v$. This gives us

$$\frac{\tan\ \theta_1}{\tan\ (-\theta_1)}=\frac{AO'}{-u}\div\frac{AO'}{v}=-\frac{u}{v} \qquad \cdots(2{\cdot}18)$$

Eliminating v/u between Eqs. (2·17) and (2·18) we get

$$n_1y_1\tan\theta_1=n_2y_2\tan\theta_2 \qquad \cdots(2{\cdot}19)$$

This is known as *Lagrange's equation for magnification.*

Eq. (2·19) is of very general application. Since longitudinal distances are not involved and only the slopes come in, it can be applied to a succession of coaxial surfaces irrespective of their separations. Let $n_1, n_2 \cdots n_p$ be refractive indices of successive media, $\theta_1, \theta_2 \cdots \theta_p$ the slopes of a given ray in these media and $y_1, y_2 \cdots y_p$ the heights of the object and successive images. Then we may apply Eq. (2·19) in succession to get

$$n_1y_1\tan\theta_1=n_py_p\tan\theta_p \qquad \cdots(2{\cdot}20)$$

This is the ***Lagrange-Helmholtz Law of Magnification.***

2·7. Relationship between f any f′ for any Optical System. Eq. (2·13), proved again as Eq. (2.16), has a general character. This can be proved by combining the Lagrange-Helmholtz law with Eqs. (2·10)—both these being general laws.

As Fig. 2·5 shows,

$$\frac{\tan\theta}{\tan\theta'}=\left(\frac{h}{-u}\right)\div\left(-\frac{h}{v}\right)=\frac{v}{u} \qquad \text{...(2·21)}$$

Eqs. (2·10*a*) and (2·10*b*) give (using $m\equiv y'/y$),

$$\frac{v}{u}=\frac{v-f'}{f}=-\frac{f'-v}{f'}\cdot\frac{f'}{f}=-\frac{y'}{y}\cdot\frac{f'}{f}$$

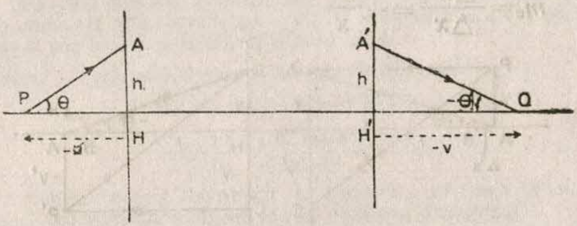


Fig. 2·5.

$$\frac{f'}{f}=-\frac{y'}{y}\cdot\frac{f'}{f} \qquad \text{...(2·22)}$$

Comparing with Eq. (2·21) we find that Gauss' general theory gives :

$$\frac{\tan\theta}{\tan\theta'}=-\frac{y'}{y}\cdot\frac{f'}{f} \qquad \text{...(2·23)}$$

But Lagrange-Helmholtz law requires that

$$\frac{\tan\theta}{\tan\theta'}=\frac{n'y'}{ny} \qquad \text{...(2·24)}$$

Comparison directly yields the result

$$\boxed{\frac{f}{f'}=-\frac{n}{n'}} \qquad \text{...(2·25)}$$

as a universal truth. Since *n* and *n′* are always positive, we see that *f′* and *f* must always have opposite signs. In the commonest case where the initial and final media are identical (being air) we get the simple result

$$\boxed{f=-f'} \qquad \text{...(2·26)}$$

2·8. Relation between Lateral, Axial and Angular Magnifications. In Fig. 2·6 let *A*, *A′* be a pair of conjugate points. At *A* we draw a transverse line of short length *AP*, an axial line of short length AB and a ray AD of slope angle θ. The conjugates of these are lines $A'P'$, $A'B'$ and ray $D'A'$ of slope θ'. By definition,

$$m = \text{lateral magnification} = \frac{y'}{y}$$

$$m_\alpha = \text{axial magnification} = \frac{\Delta x'}{\Delta x} = \frac{\Delta v}{\Delta u}$$

$$\gamma = \text{angular magnification} = \frac{\theta'}{\theta}$$

Differentiating Eq. (2·8a) we obtain

$$x.\Delta x' + \Delta x.x' = 0$$

$$m_\alpha = \frac{\Delta x'}{\Delta x} = -\frac{x'}{x}$$

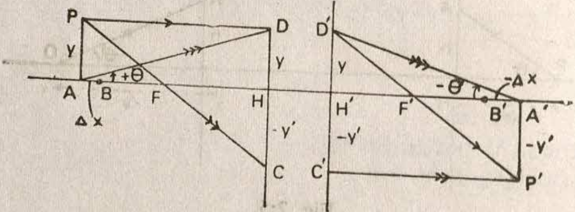


Fig. 2·6.

Showing the relations between the three magnifications m, m_α, γ.

Substituting for x' and x from Eq. (2·8b), putting $y'/y = m$,

$$m_\alpha = -\frac{-mf'}{-f/m} = -m^2 . \frac{f'}{f}$$

But $f'/f = -n'/n$, where n and n' are refractive indices of initial and final media. Hence we get

$$\boxed{m_\alpha = m^2 . \frac{n'}{n}} \qquad ...(2·27)$$

For small angles we get from Helmholtz law

$$\text{angular magnification } \gamma = \frac{\theta'}{\theta} = \frac{ny}{n'y'} = \frac{n}{n'} \cdot \frac{1}{m}$$

or

$$\boxed{\gamma m = \frac{n}{n'}} \qquad ...(2·28)$$

Finally, from Eqs. (2·27), and (2·28)

$$\boxed{\frac{m_\alpha \gamma}{m} = 1} \qquad ...(2·29)$$

These relations hold for any co-axial optical system.

For the commonest case of $n' = n$, we find that

$$m_\alpha = m^2 \qquad ...(2·30)$$

$$\gamma m = 1 \qquad ...(2·31)$$

Eq. (2·30) says that the longitudinal magnification m_α is proportional to the *square* of the lateral magnification. This is a matter of some concern. Consider a three-dimensional frame 5 mm × 5 mm transversely and 1 mm (one-fifth) longitudinally. If its image is formed with lateral magnification (say) 6, then the image would be 30 mm × 30 mm transversely and 36 mm (1·2 times) longitudinally. But no amount of ingenuity can improve the situation because Eq. (2·30) has general validity.

Again Eq. (2·31) tells that angular magnification and lateral magnifications have reciprocal relationship always. There can be no device for increasing both of them.

Example 2·4. From a given object point *A* on the axis of an image forming system, the path of a ray making 20° angle with axis was traced. The emergent ray came out with slope angle −4°. If $n=1{\cdot}62$ and $n'=1{\cdot}00$, deduce the magnification of the image of an object placed at *A*.

Solution. Lagrange Helmholtz law (Eq. 2·20) gives

$$1{\cdot}00\times y'\times\tan(-4^\circ)=1{\cdot}62\times y\times\tan 20^\circ$$

$$\therefore \qquad m=\frac{y'}{y}=\frac{1{\cdot}62\times{\cdot}3640}{-0{\cdot}0699}=-8{\cdot}4$$

[Note that immersing the object in a medium of high *n* increases the magnification. Oil immersion objective (§ 5·4) is one example.] ⊙

Example 2·5. For a co-axial image forming system the initial and final media are water ($n=1{\cdot}33$) and air respectively. For a given object the lateral magnification is −2·0. Calculate (i) the axial magnification, and (ii) angular magnification.

Solution. Given $n/n'=1{\cdot}33, \qquad m=-2{\cdot}0$

$$m_\alpha=m^2\,\frac{n'}{n}=\frac{4{\cdot}0}{1{\cdot}33}=3{\cdot}0$$

$$\gamma=\frac{n}{mn'}=\frac{1{\cdot}33}{-2{\cdot}0}=-0{\cdot}67$$ ⊙

PROBLEMS

2·1. A three dimensional object cannot be imaged in true proportions except at magnification unity. Discuss.

2·2. The Lagrange-Helmholtz law holds for large ray inclination (θ, θ'), but requires that radii of curvature of the surfaces involved are large. Comment critically.

2·3. In a given optical system, with identical media on both sides, the focal points *F* and *F′* are located at 20·7 cm and 48·3 cm respectively from an arbitrary origin. Also, for an object placed at 12·7 cm the image is formed at location 77·2 cm. Deduce :

(i) the locations of *H* and *H′*.

(ii) transverse magnification *m*.

(iii) axial magnification m_α.

2·4. In a given optical system the initial and final media have refractive indices 1·00 and 1·50 respectively. From a common origin the focal points of object space and image space are found located at 30·6 cm and 49·4 cm respectively. If an object located at −2·7 cm forms an image at 68·2 cm, deduce f, f' and the locations of H and H'.

2·5. In a system of co-axial spherical surfaces the paths of three incident rays at angles 5°0′ 10°0′, and 20° 0′ from a given axial point P are traced. The conjugate emergent rays meet the axis at three different points Q_1, Q_2, Q_3 making angles −1° 9′, −2° 18′, −4° 30′ respectively. If an object 1·00 mm tall is placed at P, and the initial and final media have refractive indices 1·65 and 1·00 respectively, deduce what Helmholtz Laws says about size of the image for rays in the three specified angular zones ?

2·6. A convergent lens, which has focal length 15·0 cm in air and whose material hasrefractive index $n=1\cdot54$, is wholly immersed in a liquid of $n=1\cdot67$. Show that it will behave as a divergent lens of focal length 104 cm.

3

Thick Lens and Lens Combinations

In this chapter we shall derive the focal lengths f, f' and positions of cardinal points H, H'; F, F' for some optical systems. 'Thick Lens' and 'two thin lenses at a distance' give us cases where the system can be treated as made of two 'thin' units. Later we shall also consider a case of combining *any* two units, which can be extended to deal with cases of any number of units.

3·1. **Thick lens in air***. In Fg. 3·1, OB and $O'C$ represent two spherical surfaces of radii of curvature R_1, R_2, separated by distance $OO'=t$ and enclosing a medium of refractive index n. This is called a 'thick lens' if t is not negligible in proportion to R_1, R_2.

To obtain the focal length and positions of cardinal points, consider an incident ray AB parallel to the principal axis. In two

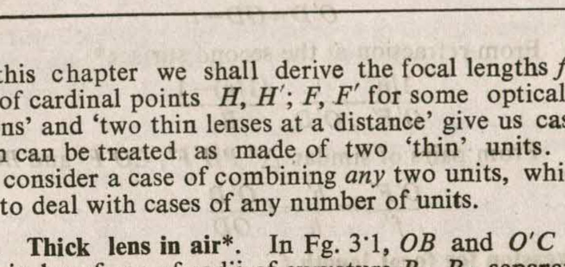


Fig. 3·1.

refractions it bends as BC, CF' meeting the axis at F'. By defintion, F' is the *focal point of image space*. If we produce AB anid CF' they meet at P. This is where the emergent ray has the same height h as the incident ray. Hence PH' gives us the *principal plane of image space*. $H'F'=f'$ is the *focal length of image space*.

Our purpose is to deduce an expression for (i) the focal length f', and (ii) the distance $O'F'$ (or $O'H'$). These two results are enough to locate F' and H'. Then we can write the expressions for f, OF and OH (not shown in Fig. 3·1) by symmetry. That will give the locations of F and H.

---

*For sign convention, see § 2·1.

The deduction is based on four fundamental relations of the case, which are as follows :

(1) For refraction at the first surfacc, using Eq. (2·11),

$$\frac{n}{OD}-\frac{1}{\infty}=\frac{n-1}{R_1}\Rightarrow OD=\frac{nR_1}{n-1} \qquad \text{...(3·1)}$$

(2) From consideration of the lens thickness,

$$O'D=OD-t \qquad \text{...(3·2)}$$

(3) From refraction at the second surface*

$$\frac{1/n}{O'F'}-\frac{1}{O'D}=\frac{(1/n)-1}{R_2} \qquad \text{...(3·3)}$$

(4) From pairs of similar $\triangle^s$ $PH'F'$, $CO'F'$ and BOD, $CO'D$

$$\frac{O'F'}{f'}=\frac{h'}{h}=\frac{O'D}{OD} \qquad \text{...(3·4)}$$

(A) Expression for focal length f'

The four equations (3·1) to (3·4) lead to

$$\boxed{\frac{1}{f'}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)+\frac{(n-1)^2}{R_1R_2}\,\frac{t}{n}} \qquad \text{...(3·5)}$$

derivation of which is left as an exercise to the reader.

(B) Expressions for distances $O'F'$ and $O'H'$

We shall express distance O' to F' by symbol β'. Eqs. (3·4), (3·2) and (3·1) lead to

$$\beta'=f'\,\frac{O'D}{OD}=f'\left(1-\frac{t}{OD}\right)$$

$$=f'\left[1-\frac{t\,(n-1)}{nR_1}\right] \qquad \text{(...3·6}a)$$

If distance O' to H' is denoted by α', we get

$$\alpha'=\beta'-f'=-f'\,\frac{t\,(n-1)}{nR_1} \qquad \text{...(3·6}b)$$

(C) Expressions for f, β and α

To get expressions for these one may consider the incident beam travelling from right to left in Fig. 3·1 and then reverse all signs, apart from interchanging R_1 and R_2. This leads to

---

*In this we presume the final medium to be the same as the initial one. Eqs. (3·5) etc. are not valid if media on two sides of the lens are different, as in Problem 3·10.

$$-\frac{1}{f}=(n-1)\left(\frac{1}{R_1}-\frac{1}{R_2}\right)+\frac{(n-1)^2}{R_1R_2}\cdot\frac{t}{n}=\frac{1}{f'}\quad\ldots(3\cdot7)$$

$$\beta=f\left[1+\frac{t(n-1)}{nR_2}\right];\quad \alpha=\beta-f=+f\,\frac{t(n-1)}{nR_2}\quad\ldots(3.8)$$

If may be noted that in the expression for α (which refers to object space distance O to H) we have R_2, not R_1. Similarly, in the expression for α' we have R_1, not R_2. Thus *it is the curvature of the 'other' end face which determines how far the principal plane would be from a given face of the lens.*

The most convenient procedure for plotting the cardinal points is to deduce α and α', which give us the principal points, and then to use f and f' values to get the focal points.

Example 3·1. Deduce the positions of cardinal points for a plano convex lens of refractive index 1·6, the radius of curvature of the curved surface being 6 cm., and thickness 2·0 cm.

Solution. Let light enter from the curved surface. We then have $R_1=+6$ cm, $t=2$ cm, $R_2=\infty$. Whence from Eqs. (3·5), (3·6) and (3·8)

$$\frac{1}{f'}=(1\cdot6-1)\left(\frac{1}{6}\right);\ \Rightarrow\ f'=\frac{6}{0\cdot6}=+10\cdot0\text{ cm.}$$

$$\alpha'=-f'\ \frac{(n-1)t}{nR_1}=-10\times\frac{0\cdot6\times2}{1\cdot6\times6}=-1\cdot25\text{ cm.}$$

$$\alpha=-f'\ \frac{(n-1)\,t}{nR_2}=0\text{ cm.}\qquad(\because\ R_2=\infty)$$

Fig. 3·2 depicts the results. First, $\alpha'=-1\cdot25$ cm, gives H' *to the left of* O'. Next use $f'=+10\cdot0$ cm, which gives F' at $10\cdot0$ cm, *to the right of* H'. Since $\alpha=0$, the point H falls on the curved surface. Finally, we have $f=-f'=-10\cdot0$ cm. So F lies $10\cdot0$ cm *to the left of* H.

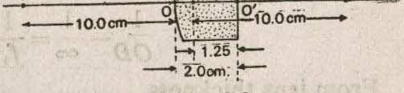


Fig. 3·2.

Check that the same results come if light enters from the right so far as the positions are concerned. Only Symbols F, H and F', H' get interchanged. ⊙

Example 3·2. A convex meniscus lens has radii of curvature 3·0 and 2 cm. and thickness 1·0 cm. If $n=1{\cdot}50$, deduce the positions of cardinal points.

Solution. *A 'meniscus' lens is one which has both surfaces curved in the same direction.* If the face on the concave side is less curved (larger R) then the lens is convex.

Fig. 3·3 shows the lens. With light travelling left to right, $R_1=-3$, $R_2=-2$, $t=1$ (all in cm.). From Eq. (3·5) we have, with $n=1{\cdot}5$.

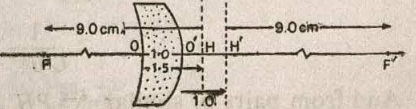


Fig. 3·3

$$\frac{1}{f'}=(0{\cdot}5)\left(\frac{1}{-3}-\frac{1}{-2}\right)+\frac{(0{\cdot}5)^2\times 1}{1{\cdot}5\,(-3)\,(-2)}$$

$$\therefore \qquad f'=+9{\cdot}0 \text{ cm. (check)}, \qquad f=-9{\cdot}0 \text{ cm.}$$

From Eq. (3·6)

$$\alpha'=-9{\cdot}0\times\frac{(1{\cdot}5-1)\times 1}{1{\cdot}5\times(-3)}=+1{\cdot}0 \text{ cm.}$$

From Eq. (3·8)

$$\alpha=-9{\cdot}0\times\frac{(1{\cdot}5-1)\times 1}{1{\cdot}5\times(1-2)}=+1{\cdot}5 \text{ cm.}$$

We plot H' at 1·0 cm *to the right of* O' and H at 1·5 cm to the right of O. Now F' lies at 9·0 cm to the right of H', and F at 9·0 cm to the left of H. Note that both the principal planes fall outside the lens and on the same side. (Check if it is a general truth about meniscus lenses). ⊙

3·2. Combination of two thin lenses. In Fig. 3·4, *OB* and *O'C* represent two "thin" lenses placed coaxially at a distance t in air. Their focal lengths are f_1' and f_2' (both for image space). We have to obtain the focal. length and positions of cardinal points of this system. The whole derivation is similar to that in § 3·1 and the labellings of Fig. 3·1 and 3·4 are identical. The four undamental equations are as follows :

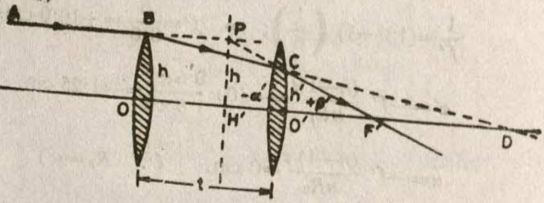


Fig. 3·4

Combination of two thin lenses

From refraction at the first lens

$$\frac{1}{OD}-\frac{1}{\infty}=\frac{1}{f_1'} \Rightarrow OD=f_1'. \qquad \ldots(3{\cdot}9)$$

From lens thickness

$$O'D=OD-t=f_1'-t \qquad \ldots(3{\cdot}10)$$

From refraction at the second lens

$$\frac{1}{O'F'}-\frac{1}{O'D}=\frac{1}{f_2'}$$

$$\therefore \qquad \frac{1}{O'F'}=\frac{f_1'+f_2'-t}{(f_1'-t)f_2'} \qquad \ldots(3{\cdot}11)$$

And from pairs of similar $\triangle^s$ $PH'F'$, $CO'F'$ and BOD, $CO'D$

$$\frac{O'F'}{f'}=\frac{h'}{h}=\frac{O'D}{OD}=\frac{f_1'-t}{f_1'} \qquad \ldots(3{\cdot}1)$$

Eqs. (3·11) and (3·12) lead to

$$\boxed{\frac{1}{f'}=\frac{f_1'+f_2'-t}{f_1'f_2'}=\frac{\triangle}{f_1'f_2'}} \qquad \ldots(3{\cdot}13)$$

where we have introduced $\triangle$ defined by

$$\triangle\equiv f_1'+f_2'-t \qquad \ldots(3{\cdot}14)$$

Also, in Eq. (3·12), $O'F'=\beta'$, whence

$$\frac{\beta'}{f'}=1-\frac{t}{f_1'}$$

$$\therefore \qquad \boxed{\alpha'=\beta'-f'=-f'\frac{t}{f_1'}} \qquad \ldots(3{\cdot}15)$$

Similarly, we may deduce

$$\boxed{\alpha=+f'\frac{t}{f_2'}} \qquad \ldots(3{\cdot}16)$$

[It may be noted that replacing $(n-1)/R_1$, $(n-1)/R_2$ and t by f_1', f_2' and t/n respectively leads from Eqs. (3·5) and (3·6) to Eqs. (3·13) and (3·15) respectively. See why it is so.]

Example 3·3. Two convergent lenses of focal lengths 20 and 10 cm are placed 25 cm apart in air. Find and plot the positions of H, H', F and F'. Also find the position and magnification of the image of an object placed 10 cm in front of the first lens.

Solution. $\triangle=f_1'-f_2'-t=20+10-25=5$ cm

$$\therefore \qquad f'=\frac{f_1'f_2'}{\triangle}=\frac{20\times 10}{5}=40 \text{ cm} ;$$

$$f=-f'=-40 \text{ cm}$$

$$\alpha'=-f'\frac{t}{f_1'}=-50 \text{ cm}$$

$$\alpha=f'\frac{t}{f_2'}=100 \text{ cm.}$$

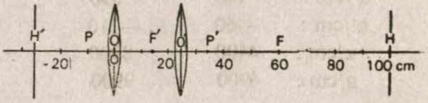


Fig. 3·5.

The cardinal points are plotted in Fig. 3·5 using these results. O to $O'=25$ cm ; O to $H=\alpha=100$ cm ; O' to $H=\alpha'=-50$ cm ; H to $F=f=-40$ cm ; H' to $F'=f'=40$ cm.

Given O to object $P=-10$ cm, we get $u=H$ to $P=-110$ cm. Now

$$\frac{1}{v}-\frac{1}{-110}=\frac{1}{40} \qquad \Rightarrow \qquad v=\frac{440}{7}=63 \text{ cm } (H' \text{ to } P')$$

That locates P', 13 cm to the right of O'. Finally

$$m=\frac{v}{u}=\frac{440}{7}\times\frac{1}{-110}=-\frac{4}{7}=-0{\cdot}57$$

One could also use Newton's method (Eqs. 2·8) :

$$x=F \text{ to } P=-70 \text{ cm} \; ; \; x'=\frac{(40)(-40)}{-70}=23 \text{ cm}=F' \text{ to } P'.$$

$$m=-\frac{f}{x}=-\frac{-40}{-70}=-\frac{4}{7}.$$

[Note that H and H' fall far from the lenses O and O' in this case. H falling far to the right makes $u=-110$ cm, though O to $P=-10$ cm only. Repeat the exercise with t=35 cm to get further insight.] ⊙

3·3. Some Special Cases of Lens Combinations.

3·3. Some Special Cases of Lens Combinations. Although there is a very large variety of combinations of lenses in optical instruments, we shall discuss only two cases along with typical numerical examples. (Eyepieces will be discussed in Chapter 6).

(A) Case of a Telescopic Combination

In this case the spacing (t) between the lenses is made equal to $f_1'+f_2'$. As a result $\triangle$ becomes zero, so that $f'=\infty$ and the cardinal points H, H' lie at infinity. It is useful to see what happens as t passes from below $f_1'+f_2'$ to above that value.

Example 3·4. Two thin lenses of focal lengths 100 cm and —10 cm form a telescope.* Discuss the variations of the range of focussing as the separation of lenses is changed about that for viewing objects at infinite distance with the relaxed eye.

Solution. The separation t, for focussing objects at ∞ with relaxed eye needs $f'=\infty$, whence $\triangle=0$, and $t=f_1'+f_2'=90$ cm. Let the actual separation be varied from 88 to 92 cm in steps of 1 cm. We shall only give the results (check them) :

t/cm :	88	89	90	91	92
$\triangle$/cm :	2	1	0	—1	—2
f'/cm :	—500	—1000	∞	+1000	+500
α'/cm :	440	890	∞	—910	—460
β'/cm :	—60	—110	∞	90	40
α/cm :	4400	8900	∞	—9800	—4600
β/cm :	4900	9900	∞	—10100	—5100

The focal plane F of object space is that for which the image is formed at ∞ and is therefore seen by the relaxed eye. The position of this with respect to the objective is given by β. For $t=88$ cm this plane is 49 metres *behind* the observer ; as t approaches 90 cm, F recedes to $-\infty$; as t just crosses 90 cm, F becomes $+\infty$ and thereafter comes closer on the *front* side, reaching 51 m as t reaches 92 cm. Since real objects only in *front* can be seen, the range of focussing is ∞ to 51 m if t is varied from 90 cm to 92 cm. ⊙

(B) Case of the Telephoto Lens

For distant objects a single lens camera of manageable length produces photographs which are quite reduced in size. A telephoto lens combination aims at obtaining photographs of distant objects reduced by a much smaller factor than what may be had from a single lens camera. This is done by using a *convex* lens of focal length

---

*This is a Gaussian telescope, using a concave lens as eyepiece.

f_1' backed by a larger power *concave* lens of focal length f_2', the separation being kept somewhat *greater* than $f_1'+f_2'$. The function is best explained with a concrete example.

Example 3·5. A camera objective has a positive lens of focal length 30·0 cm backed by a negative lens of focal length 6·0 cm at 25·0 cm distance. Deduce the position of the photographic plate, total length of the camera, and magnification of the image for an object 200 m in front.

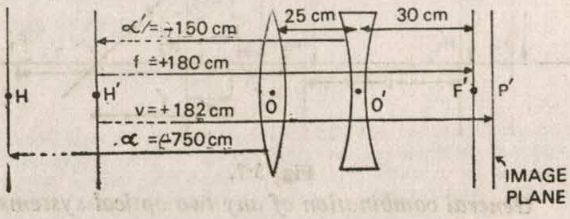


Fig. 3·6,

A telephoto lens combination has H' far towards the object to keep $O'P'$ small though v ($H'P'$) is large.

Solution. $f_1'=30{\cdot}0$ cm, $f_2'=-6{\cdot}0$ cm, $f_1'+f_2'=24{\cdot}0$ cm

$$t=25{\cdot}0 \text{ cm} \quad \Rightarrow \quad \Delta=-1{\cdot}0 \text{ cm}.$$

These lead to

$$f'=180 \text{ cm}, \quad \alpha'=-150 \text{ cm}, \quad \beta'=30 \text{ cm}$$

which locate F' ($O'F'=30$ cm) and H' ($O'H'=-150$ cm). Similarly,

$$f=-180 \text{ cm}, \quad \alpha=-750 \text{ cm}, \quad \beta=-930 \text{ cm}$$

which locates H($OH=-750$ cm), F being not shown in the figure.

Now, given P (not shown) at 200 m to the left of O, we get

$$u=HP=-200 \text{ m}-(-7{\cdot}5 \text{ m})=-192{\cdot}5 \text{ m}$$

$$\therefore \quad v=\frac{uf'}{u+f'}=1{\cdot}82 \text{ m (check it)}$$

$$m=\frac{v}{u}=\frac{1{\cdot}82}{-192{\cdot}5}\approx-\frac{1}{100}$$

Since $v=H'P'=1{\cdot}82$ m, and $H'O'=-\alpha'=1{\cdot}50$ m, we get

$$O'P'=32 \text{ cm}$$

That gives length of the camera $\approx OP'\approx$ **57 cm.** ⊙

Note that the arrangement "throws" H and H' far out towards the object. Therefore, while the focal length f' is as large as 180 cm, the second focus F' is only 30 cm behind O'. A single lens of focal length $f'=180$ cm would give about the same magnification ($\approx-1/100$), but the camera would have to be about 2 metres long. On the other hand, a camera of length 57 cm using a single lens will need a lens of $f'\approx55$ cm, and its magnification would be $\approx 55/20000\approx-1/360$, smaller by a factor of 3·6.

3·4. General Combination of two Optical Systems or of two Thick Lenses. We now consider the combination of *any* two coaxial optical systems.

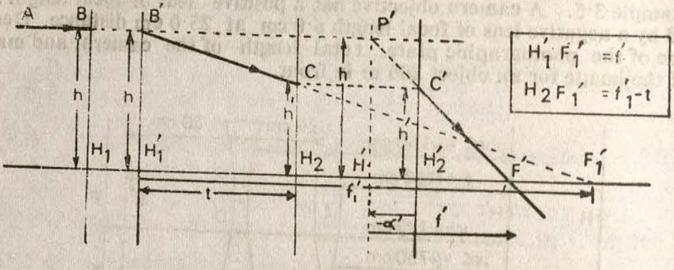


Fig. 3·7.

General combination of any two optical systems

On Fig. 3·7, the first system is represented by principal planes H_1, H_1' and the second focal point F_1' (F_1 not needed). A ray *AB* parallel to the axis at height *h* meets H_1 plane at *B*. Its conjugate ray must meet H_1' plane at the same height *h* (that gives *B′*) and must pass through F_1'. That gives $B'F_1'$ as the ray emergent from the first system.

The second system is represented by principal planes H_2, H_2'. The distance H_1' to H_2 is defined as the separation *t* of the two systems*. For this system F_1' is the object point, if the conjugate image point is *F′*, then that is the second focal point of the combined system.

Ray B' F_1' meets the plane H_2 at *C*, which has height *h′*. The conjugate emergent ray must meet H_2' plane at the same height *h′* and pass through *F′*. That gives us the final emergent ray *C′F′*.

If both *AB* and *C′F′* are produced, they meet at *P′*, which gives us plane *P′H′*, the principal plane of image space for the combined system. $H'F'=f'$ is the focal length of image space for the combined system.

We have to locate *H′* (by deducing $H_2'H'=\alpha'$) and deduce f' (so as to locate *F′*).

Now, from pairs of similar Δˢ $B'F_1'H_1'$, $CF_1'H_2$ and $P'F'H'$, $C'F'H_2'$, we get

$$\frac{h}{h'}=\frac{f_1'}{f_1'-t}=\frac{f'}{f'+\alpha'} \qquad \ldots(3{\cdot}17)$$

For refraction at the second lens, with $H_2'F'$ as *v* and H_2F_1' as *u*,

$$\frac{1}{f'+\alpha'}-\frac{1}{f_1'-t}=\frac{1}{f_2'} \qquad \ldots(3{\cdot}18)$$

---

*Notice that *t* defined this way alone leads to simple results. It could be negative if H_2 occurs earlier than H_1' on the path of light.

Eliminating α' between these gives

$$f'=\frac{f_1'f_2'}{f_1'+f_2'-t}=\frac{f_1'f_2'}{\Delta} \qquad \ldots(3{\cdot}19)$$

and solving for α' gives

$$\alpha'=-f'\ \frac{t}{f_1'}=-f_2'\ \frac{t}{\Delta} \qquad \ldots(3{\cdot}20)$$

Similarly, one can deduce

$$f=-f' \text{ and } \alpha=f'\ \frac{t}{f_2'}=f_1'\ \frac{t}{\Delta} \qquad \ldots(3{\cdot}21)$$

It is to be noted that the formulae are exactly the same as for a combination of two thin separated lenses. It means that *we can combine any number of co-axial systems into a single equivalent system* and deduce its H, H', f, f' if those for the components are known, along with the separations. This is what was proved in a more general way by Gauss (see Chapter 2).

3·5. Deviation Method for Locating the Cardinal Points. We have so far dealt primarily in terms of the axial distances. But the business of a lens is to cause angular deviation in the rays falling on it. We shall now see how through computation of deviations we can get to the same results as before. Of course, some additional insight is also obtained in the process.

In Fig. 3·8 we consider an incident ray PA and its conjugate ray $A'P'$. There is a deviation δ, and it is given (assuming small angles) by

$$\delta=\angle APP'+\angle A'P'P$$

$$=\frac{h}{-u}+\frac{h}{v}$$

$$=h\left(\frac{1}{v}-\frac{1}{u}\right)$$

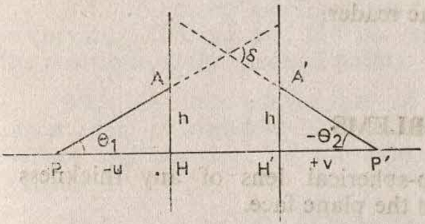


Fig. 3·8,
Deviation in a ray is h/f' for any system

or

$$\boxed{\delta=\frac{h}{f'}} \qquad \ldots(3{\cdot}22)$$

This is a remarkably simple result. The lens causes a deviation which is independent of u, and it is proportional to the height where the ray meets the lens (or the principal plane) and inversely proportional to f'. δ is taken as $+ve$ when the deviation is towards the axis.

Now consider Fig. 3·9 which corresponds to Fig. 3·7. With incident ray parallel to the axis, the first system gives deviation

$$\delta_1=h/f_1',$$

the second system gives deviation

$$\delta_2 = h'/f_2',$$

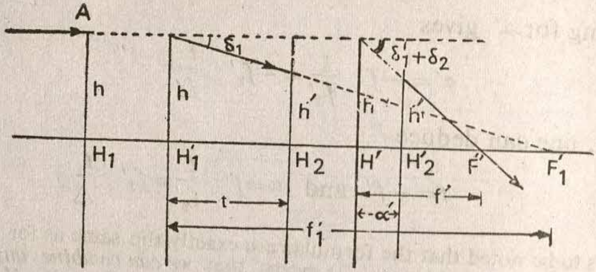


Fig. 3·9.

For a combination of two systems the deviations add together and the whole system gives deviation

$$\delta = h/f',$$

where f_1', f_2' and f' are focal lengths of the first system, the second system and the combination. Then

$$\delta = \frac{h}{f'} = \delta_1 + \delta_2 = \frac{h}{f_1'} + \frac{h'}{f_2'} \qquad \ldots(3{\cdot}23)$$

From similar Δ^s in Fig. 3·9. the relation between h and h' may be obtained. Then Eq. (3·23) leads to relations (3·13) and (3·15). The exercise is left for the reader.

PROBLEMS

3·1. Show that for a plano-spherical lens of any thickness one principal plane always lies on the plane face.

3·2. Show that for a meniscus lens the principal planes always fall outside the lens thickness and on the same side.

3·3. Show that in a combination of two coaxial lenses the distance between the two principal planes is given by t^2/Δ.

3·4. Show that in a combination of two thin lenses L_1 and L_2, with the usual notation,

$$\frac{L_1H}{L_2H'} = -\frac{f_1'}{f_2'},$$

which for a thick lens (with faces marked O_1, O_2) gives

$$\frac{O_1H}{O_2H'} = +\frac{R_1}{R_2}.$$

3·5. For a thick lens with $R_1=+10{\cdot}0$ cm, $R_2=-6{\cdot}0$ cm, $n=1{\cdot}50$ deduce the focal length f' as the separation t changes from nearly 0 to 32 cm in steps of 8 cm.

3·6. Examine whether a combination of two concave lenses can behave as a convex lens. (The reverse is known to be true.)

3·7. The numerical radii of curvature of a thick double convex lens are r_1 and r_2. Show that the lens becomes concave if $t(n-1)>n(r_1+r_2)$, where t is lens thickness and n is refractive index of the lens material.

3·8. A cylinder of transparent medium wlth radius R and refractive index n focusses rays in a section perpendicular to the axis. Deduce an expression for its focal length. What will the form of image be if a point source is used ?

3·9. Deduce the focal length and positions of cardinal points for a glass sphere ($n=1{\cdot}60$) of radius 2·0 cm placed in air. (Assume the rays being limited to the paraxial region.)

3·10. Deduce the positions of cardinal points for a hemisphere of glass ($n=1{\cdot}5$) of radius 2·0 cm placed with the plane face in contact with water ($n=1{\cdot}30$).

3·11. A double convex lens of refractive index 1·50 has radii of curvature 10·0 cm and 15·0 cm and thickness 5·0 cm. Deduce and plot the positions of the cardinal points.

3·12. A convex meniscus lens of refractive index 1·50 has radii of curvature 10·0 cm and 15·0 cm and thickness 2·0 cm. Deduce the positions of the cardinal points.

3·13. A thin convex lens of focal length 20 cm and a thin concave lens of focal length 10 cm are separated by a distance 8 cm. Deduce the positions of the principal and focal points of the system.

3·14. For the data of Problem 3·13 deduce the magnification if the object is placed 50 cm in front of the *second* lens.

3·15. Design a combination of two thin lenses for which H and H' coincide with lenses L_2 and L_1 respectively, and $f'=6$ cm.

3·16. Use Eq. (3·22) to define f' as h/δ. For the case of refraction at a single spherical surface deduce an expression for δ and hence for f'.

3·17. Two convex meniscus lenses are set coaxially with their concave faces outwards and at distance 7·0 cm. The radii of curvatures are 4·0 and 5·0 for the first lens and 6·0 and 8·0 cm for the second. The thickness is 2·0 cm for each lens, and the material for both has $n=1{\cdot}60$. Deduce the focal length of the combination. [Note that t comes out $-2\ 6$ cm, since H_2 falls to the left of H_1' with rays taken incident from the left.]

4

Dispersion of Light and Chromatic Aberration

4·1. Dispersion. The variation of refractive index n of a given medium with wavelength λ is called dispersion. If n varies rapidly with wavelength, the medium is said to be highly dispersive. In this sense, *flint glass is more dispersive than crown glass.*

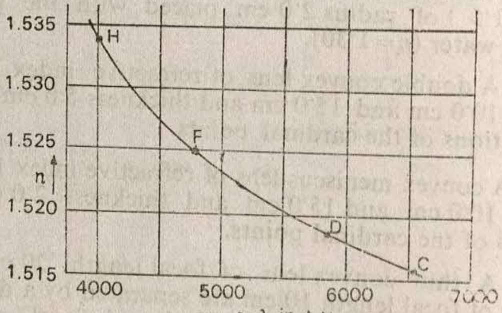


Fig. 4·1.

Dispersion curve for crown glass

Quantitative definitions for **dispersive power** are given variously. In wave theory it is defined as

$$\text{dispersive power} \equiv \frac{\text{change in } n}{\text{change in } \lambda} = \frac{dn}{d\lambda} \qquad \ldots(4{\cdot}1)$$

For light it is always a negative quantity, which means that n *decreases* as λ *increases.* Further, the relation between n and λ is not linear ; that is, $dn/d\lambda$ is not constant over a wide range of λ. Fig. 4·1 shows the general nature of the dispersion curve (n *vs.* λ). It shows that *the magnitude of* $dn/d\lambda$ *decreases as* λ *increases.*

In geometrical optics a more useful quantity is defined thus :

$$\text{dispersive power, } \omega \equiv \frac{|\, n_2 - n_1 \,|}{n-1} \qquad \ldots(4{\cdot}2)$$

where n_2 and n_1 are the refractive indices for *two suitably chosen colours* and n that for a *mean colour* between 1 and 2. There is no universal choice about colours 1 and 2 chosen for defining ω by Eq. (4·2). Reason : the choice depends upon the range of colours within which one has to work. For vision in day-light the range blue to red is most useful, while for usual photography the range violet to green is most useful. Obviously, the optician and photographer use different colour limits to define ω of the materials of their lenses.

The optician's precise definition of dispersive power is

$$\omega=\frac{n_F-n_C}{n_D-1}, \qquad \ldots(4·3)$$

in which the suffixes refer to light of definite wavelengths* :

$C(\lambda=6563\ A°$, red) α line of H spectrum

$D(\lambda=5893\ A°$, yellow) yellow line of Na spectrum

$F(\lambda=4861\ A°$, blue) β line of H spectrum

In Table 4·1 we give n_C, n_D, n_F and ω values of some of the important glasses. It will be noted that flint is one-and-a-half times dispersive compared to crown, while *DEDF* is more than two-fold dispersive.

TABLE 4·1

n and ω values of some standard optical glasses

Name	n_C $\lambda=6563\ A°$	n_D $\lambda=5893\ A°$	n_F $\lambda=4861\ A°$	n_H $\lambda=4047\ A°$	ω
Hard Crown	1·5164	1·5190	1·5249	1·5334	0·0166
Flint	1·5744	1·5786	1·5885	1·6035	0·0243
EDF*	1·6936	1·7003	1·7168	1·7427	0·0330
DEDF**	1·7407	1·7484	1·7675	1·7981	0·0360

*Extra dense flint. **Double extra dense flint.

4·2. Longitudinal Chromatic Aberration of an Optical System. For a reflecting system the focal length depends on purely geometrical factors and hence for a given object point rays of all colours get focussed at the same point. But *in a refracting system rays of*

---

[*] Designations $A, B, C, D,\ldots F\ldots$ were given by Fraunhofer in the absorption lines seen in the solar spectrum, starting from the long wavelength end.

different colours from a given object point give different image points. This 'defect' is called chromatic aberration.

In Fig. 4·2 the action of a convex and a concave lens on a parallel beam of white light is shown. In each case, the violet rays come to focus nearer to the lens than the red ones.

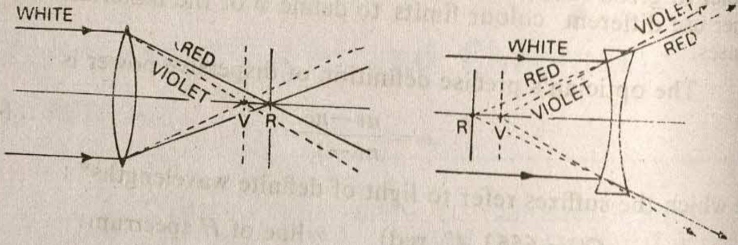


Fig. 4·2.

Longitudinal chromatic aberration

The quantity $f'_C - f'_F = \Delta f'$ is defined as the longitudinal chromatic aberration of an optical system, where C and F refer to Fraunhofer lines 6563 $A°$ (red region) and 4861 $A°$ (blue region) respectively.*

To deduce the aberration $\Delta f'$ we write the focal length equation for a thin lens as

$$\frac{1}{f'}=(n-1)\ G \qquad \text{...(4·4)}$$

where G is the geometrical factor $(1/R_1)-(1/R_2)$. Now differentiating with respect to colour gives

$$\Delta\left(\frac{1}{f'}\right)=G.\Delta n=\frac{1}{f'}\cdot\frac{\Delta n}{n-1}=\frac{\omega}{f'} \qquad \text{...(4·5)}$$

Hence,

$$|\ \Delta f'\ | =\omega f' \qquad \text{...(4·6)}$$

One may choose the appropriate colours to specify ω and use focal length of the mean colour as f' here.

For a point source at finite distance, the longitudinal aberration is

$$\Delta v = v_C - v_F.$$

Since $\frac{1}{v}-\frac{1}{u}=\frac{1}{f'}$, we get

---

*This choice suits visual observations. The photographer would prefer the choice of F (4861 $A°$) and G (4358 $A°$) lines.

$$\Delta v = v^2 \frac{\Delta f'}{f'^2} = \omega \frac{v^2}{f'}$$
$$= \frac{v^2}{f'^2} \cdot |\Delta f'| \qquad \ldots(4{\cdot}7)$$

It will be noted from Fig. 4·2 (left) that if a screen is placed at V, the image has a violet centre with other colours surrounding it, while if the screen is placed at R the image has a red centre with violet outermost. The *best compromise* image is obtained in some intermediate position of the screen.

Example 4·1. A lens made of crown glass has focal length 50 cm. Calculate the longitudinal chromatic aberration given, $n_C = 1{\cdot}5164$, $n_F = 1{\cdot}5249$.

Solution. n_D being not given, we use mean $n = 1{\cdot}5206$ for it.

$$\omega = \frac{1{\cdot}5249 - 1{\cdot}5164}{{\cdot}5206} = \frac{{\cdot}0085}{{\cdot}5206} = 0{\cdot}0164$$

Longitudinal chromatic aberration $x = f' \ \omega = 50 \times {\cdot}0164 =$ **8·82 cm.** ⊙

Example 4·2. The lens of Example 4·1 is used to form the image of a point source placed 70 cm in front. Deduce the longitudinal focussing difference for C and F light.

Solution. One can easily get v=175cm. Then using Eq. (4·7), and $\omega =$ ·0164 obtained previously,

$$\Delta v = {\cdot}0164 \times \frac{(175)^2}{50} = \mathbf{10{\cdot}0\ cm.}$$

[Notice the largeness of the error. In daylight we do not have much intensity in F region. So the observed aberration is between C and D light, which is smaller.] ⊙

4·3. Lateral Chromatic Aberration. In Fig. 4·3 an object of height y is shown imaged by a lens for rays of different colours corresponding to C and F lines of Fraunhofer. We see that the images

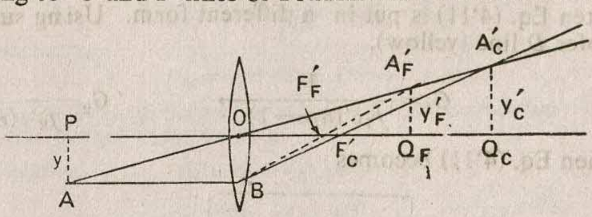


Fig. 4·3.

Lateral chromatic aberration

A_F', A_C' not only fall in different planes, but also fall at *different heights* y_F', y_C' from the axis. The difference of magnification $\Delta m = (y_C'/y) - (y'_F/y)$ is defined as the lateral chromatic aberration. However, it may be more appropriate to define the *ratio* $(\Delta m)/m$ as the lateral chromatic aberration. It can be shown that

$$\Delta m = \Delta f' \frac{u}{(u+f')^2}; \qquad \frac{\Delta m}{m} = \omega \cdot \frac{u}{u+f'} \qquad \ldots(4{\cdot}8)$$

4·4. Achromatic Combination of Thin Lenses in Contact. We may choose and arrange the components of an optical system in such a way that *for rays of two different colours a given object point gives a single image point*. The system is then called achromatic (*i.e.*, free from chromatic aberration) *for the chosen pair of colours.* We shall consider the case of two thin lenses in contact. Let the lenses have constants $n_{1\mathrm{C}}$, $n_{2\mathrm{F}}$, G_1 and $n_{2\mathrm{C}}$, $n_{2\mathrm{F}}$, G_2. Then for Fraunhofer C line the focal length of the combination in contact is given by

$$\frac{1}{f'_{\mathrm{C}}}=(n_{1\mathrm{C}}-1)\,G_1+(n_{2\mathrm{C}}-1)\,G_2. \qquad \ldots(4{\cdot}9)$$

and similarly for Fraunhofer F line

$$\frac{1}{f'_{\mathrm{F}}}=(n_{1\mathrm{F}}-1)\,G_1+(n_{2\mathrm{F}}-1)\,G_2 \qquad (\ldots4{\cdot}10)$$

For achromatism $f'_{\mathrm{C}}=f'_{\mathrm{F}}$. Hence

$$\boxed{(n_{1\mathrm{F}}-n_{1\mathrm{C}})\,G_1+(n_{2\mathrm{F}}-n_{2\mathrm{C}})\,G_2=0} \qquad \ldots(4{\cdot}11)$$

Equation (4·11) is the desired condition for achromatism. Along with (4·9) or (4·10) it enables us to design an achromatic doublet. We are given the four refractive indices and the *desired* focal length ($f'_{\mathrm{C}}=f'_{\mathrm{F}}$), from which we deduce the geometric factors for the component lenses (G_1 and G_2).

Some freedom is still available, since with a given G we can choose the radii R and R' in many ways. The usual practice is to keep one face plane and the radii of curvature of the contacting faces of the two lenses equal.

Often Eq. (4·11) is put in a different form. Using suffix D for Fraunhofer D line (yellow),

$$G_1=\frac{1}{f_{1\mathrm{D}}'(n_{1\mathrm{D}}-1)}, \qquad G_2=\frac{1}{f_2'{}_{\mathrm{D}}(n_{2\mathrm{D}}-1)}$$

Then Eq. (4·11) becomes

$$\boxed{\frac{\omega_1}{f_1'{}_{\mathrm{D}}}+\frac{\omega_2}{f_2'{}_{\mathrm{D}}}=0} \qquad \ldots(4{\cdot}12)$$

where ω_1 and ω_2 are dispersive powers of materials of the two lenses. If ω_1, ω_2 and f' of the combination are given, Eq. (4·12) with

$$\frac{1}{f_1'{}_{\mathrm{D}}}=\frac{1}{f_1'{}_{\mathrm{D}}}+\frac{1}{f_2'{}_{\mathrm{D}}} \qquad \ldots(4{\cdot}13)$$

enables us to 'design' an achromatic doublet. It is achromatic for colours *in the neighbourhood of D.**

Example 4·3. Dispersive powers of Crown and Flint glasses are ·0163 and ·0243 respectively. Design an achromatic contact doublet of focal length 50 cm.

Solution. Straight substitution in Eqs. (4·12) and (4·13) gives

$$\frac{\cdot 0163}{f_1'}+\frac{\cdot 0243}{f_2'}=0\,; \qquad \frac{1}{f_1'}+\frac{1}{f_2'}=\frac{1}{50}$$

Solving, we get $f_1'=\dfrac{4000}{243}=16{\cdot}5$ cm. (*crown*)

$$f_2'=\frac{4000}{-163}=-24{\cdot}5 \text{ cm. (\textit{flint})}$$ ⊙

Example 4·4. A contact achromat of power +1 is to be made using crown and flint glass lenses. For crown $n_C=1{\cdot}51645$, $n_F=1{\cdot}52492$. For flint $n_C=1{\cdot}57446$, $n_F=1{\cdot}58854$. Calculate the focal lengths of the component lenses.

Solution. Nature of data desire using Eqs. (4·11) and (4·9). Now,

Using suffixes 1 and 2 for crown and flint respectively, we have

$$n_{1_F}-n_{1_C}=0{\cdot}00847, \quad n_{2_F}-n_{2_C}=0{\cdot}01408$$

For achromatism, Eq. (4·11) gives

$$0{\cdot}00847\ G_1+0{\cdot}01408\ G_2=0$$

For the combination power $+1, f'=100$ cm. Hence from Eq. (4·9).

$$0{\cdot}5164\ G_1\ +0{\cdot}5745\ G_2=\frac{1}{100}$$

We have to solve these two equations for G_1 and G_2. This gives

$$G_1=0{\cdot}0585. \qquad G_2=0{\cdot}0352 \text{ (check it)}$$

Finally,

$$\frac{1}{f_1'{}_C}=0{\cdot}5164\ G_1=0{\cdot}5164\times0{\cdot}0585$$

$$\therefore \qquad f_1'{}_C=33{\cdot}1 \text{ cm (\textit{crown})}$$

Similarly,

$$f'_{2C}=-49{\cdot}4 \text{ cm. (\textit{flint})}$$ ⊙

Example 4·5. In the doublet of Ex. 4·4 if the contact faces are both plane, complete the 'design' of the doublet.

Solution. $R_1'=R_2=\infty$. Hence

$$\frac{1}{R_1}=G_1 \quad \text{and} \quad -\frac{1}{R_2'}=G_2$$

Using the values of G_1 and G_2 deduced above, we get

$$R_1=17{\cdot}1 \text{ cm}, \qquad R_2'=28{\cdot}4 \text{ cm.}$$ ⊙

4·5. Achromatic Combination of Two Separated Lenses. For a combination of two coaxial thin lenses separated in air by distance t, Eq. 3·13 gives

$$\frac{1}{f'}=\frac{1}{f_1'}+\frac{1}{f_2'}-\frac{t}{f_1'f_2'} \qquad \ldots(4{\cdot}14)$$

---

*This procedure is used if ω values are given. If n_C and n_F values are given, one must use Eqs. (4·11) and (4·9).

If we differentiate this with respect to colour and express the differences by Δ, then for achromatism ($\Delta f'=0$) we need

$$\Delta\left(\frac{1}{f_1'}\right)+\Delta\left(\frac{1}{f_2'}\right)-\frac{t}{f_1'}\Delta\left(\frac{1}{f_2'}\right)-\frac{t}{f_2'}\Delta\left(\frac{1}{f_1'}\right)=0 \qquad \ldots(4{\cdot}15)$$

Now,

$$\Delta\left(\frac{1}{f_1'}\right)=\Delta n_1.G_1=\frac{\Delta n_1}{n_1-1}\,\frac{1}{f_1'}=\frac{\omega_1}{f_1'}, \qquad \ldots(4{\cdot}16)$$

where ω_1 is dispersive power of the material of the first lens and G is the geometric factor $(1/R_1-1/R_2)$. Using a similar expression for $\Delta\,(1/f_2')$ and substituting in Eq. (4·15) we get for achromatism

$$\boxed{\frac{\omega_1}{f_1'}+\frac{\omega_2}{f_2'}-\frac{t}{f_1'f_2'}(\omega_1+\omega_2)=0} \qquad \ldots(4{\cdot}17)$$

If the lenses are of identical materials ($\omega_1=\omega_2$), the condition for achromatism reduces to

$$\boxed{t=\frac{f_1'+f_2'}{2}} \qquad \ldots(4{\cdot}18)$$

Thus an achromatic combination of two thin lenses of the same material is achieved if separation between them is equal to half the sum of their focal lengths.

It is interesting to recognize that since f_1' and f_2' depend on λ we can satisfy Equation (4·18) for one particular wavelength λ_0 only. This means that *in the neighbourhood of this wavelength* the focal length does not change with λ. This λ_0 may be chosen at sodium yellow (5893 $A°$) for visual instruments. Then f'_C and f'_F will be nearly, but not rigorously, equal.

The contact achromat is widely used in objectives of telescopes, collimators and cameras. But in eyepieces the separated lens achromat is widely used (see Chap. 6).

4·6. Achromatism of size, not of position of images. It is to be noted that what is achieved from Equation (4·18) is *not the coincidence of focal points, but equality of focal lengths for different colours*. To understand this, we note that distance α' of the second principal plane of the combination from the second lens is given for Fraunhofer C and F lines [vide Eq. 3·15] by

$$\alpha_C'=-f_C'\,\frac{t}{f_{1C}'} \quad \text{and} \quad \alpha_F'=f_F'\,\frac{t}{f_{1F}'}$$

Now in achromatism we make $f_C' = f_F'$, but f_{1C}' is not equal to f_{1F}'. Hence the combination *does not have its principal planes* H_C' *and* H_F' *in coincidence*. The focal points lie at distance $f_C' = f_F'$ from the respective principal planes. Hence the focal points also do not coincide. Fig. 4·4 shows the principal and focal planes of the image space (those for object space not shown). For a white incident ray the emergent rays for C and F colours are as shown. They *make equal angles with the axis but do not meet the axis at the same point*.

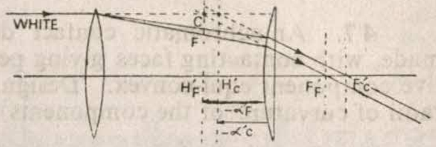


Fig. 4·4.

An optician does not mind this, because vision depends upon angles subtended. But a photographer will reject the combination.

A hard fact is that no combination of lenses of the same material can achieve achromatism of both position and magnification of the image. We shall not proceed to give a formal proof of this.

PROBLEMS

4·1. Dispersive power of a material is sometimes defined as $(n_{violet} - n_{red})/(n_{yellaw} - 1)$. Discuss the need for a more precise definition. Explain also why an optician and a photographer require different reference 'colours' for defining dispersive power.

4·2. Using values of Table 4·1, calculate the longitudinal chromatic aberration for a convergent lens of crown glass of focal length 1·0 metre, object being considered (*i*) at infinity, (*ii*) at 2 metres in front of the lens.

4·3. Spectacle lenses are generally made of crown glass, while spectrometer prisms are made of flint glass. Explain.

4·4. Chromatic 'error' of a lens does not depend on the aperture. But if the aperture of a lens is reduced, the chromatic *defect* in the image on a screen is also reduced. Explain.

4·5. An achromatic contact doublet of focal length 40·0 cm is to be made using hard crown and flint glasses. Calculate the focal lengths of the components from the following data :

Hard crown $n_D = 1{\cdot}5190$ $n_F - n_C = 0{\cdot}00852$

Flint $n_D = 1{\cdot}5786$ $n_F - n_C = 0{\cdot}0141$

Calculate also the geometric factors $[G = (1/R - 1/R')]$ for the two lenses.

4·6. From the data in Table 4·1, design contact doublets of $f'=100$ cm using Hard Crown and *EDF* glasses and achromatic fro (*i*) *C* and *D*, (*ii*) *F* and *H*.

4·7. An achromatic contact doublet of $f'=80{\cdot}0$ cm is to be made, with contacting faces giving perfect fit, and keeping the positive component equi-convex. Design the doublet (*i.e.*, deduce the radii of curvature of the components), given the following data :

for crown glass : $n_C=1{\cdot}51642$, $n_F=1{\cdot}52494$

for flint glass : $n_C=1{\cdot}57443$, $n_F=1{\cdot}58852$

4·8. Show that for an achromatic combination of two separated lenses of the same material

$$\frac{1}{f'}=\frac{1}{2}\left(\frac{1}{f_1'}+\frac{1}{f_2'}\right)$$

4·9. Discuss the limitations of an achromatic combination of two thin lenses placed at a distance apart, and specifically answer whether this combination is suitable for (*i*) a camera objective, (*ii*) an eyepiece, (*iii*) a telescope objective.

4·10. From the data in Table 4·1, deduce the angular separation of *C* and *F* lines by a 60° prism in minimum deviation position, if the prism is made of (*i*) hard crown glass, (*ii*) *DED* Flint glass.

4·11. A single lens magnifier has focal length 5·0 cm for *D* line, and is used as a×5 magnifier for that colour. If dispersion from *C* to *F* is 0·017, deduce the lateral chromatic aberration $|m_C-m_F|$.

5

Monochromatic Aberrations of Images

In this chapter we shall consider aberrations present in images formed with monochromatic light. These aberrations occur if (*a*) the optical system has large aperture, or (*b*) the rays make large angles with the principal axis, or (*c*) the object has large size perpendicular to the axis.

In the ideal theory of image formation we always use the approximation $\sin\theta=\theta$. This is not true if the angles involved are large. We may get a first estimate of the errors if we take $\sin\theta=\theta-\frac{\theta^3}{3!}$. The theory developed is then called the '*third order*' *theory*. For even greater accuracy θ^5 term may also be considered (*fifth order theory*). We shall not give these theories, but only refer to the results and indicate the steps needed to avoid or minimise the aberrations.

5·1. General Description of Different Aberrations

(*i*) **Spherical Aberration.** If we consider a lens with *spherical surfaces*, and an object point *P on the axis*, then the rays near the

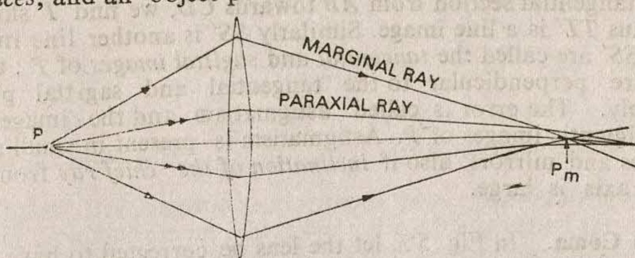


Fig. 5·1.

Longitudinal spherical aberration for object at finite distance

axis (*paraxial rays*) and the rays near the margin of the lens (*marginal rays*) form images at different points P_a and P_m [Fig. 5·1]. The distance P_aP_m is called *longitudinal spherical aberration.* In a transverse plane at the paraxial image point P_a the image is not a point, but is a circle of some radius. This radius gives the *transverse* *spherical aberration.* Spherical aberration is due to *large aperture* of the lens* and concerns the image of an axial object point.

(*ii*) **Astigmatism.** In Fig. 5·2, P is a point object far aside from the axis of the lens. The plane passing through P and the principal axis is called the *tangential* or *meridianal section.* The plane perpendicular to this and passing through the principal axis is called the *sagittal section.* Now we consider only a *small aperture ABCD of the lens near the centre.* Even then, due to *large inclination* of the chief ray with the axis we find that a narrow pencil of rays from P does not meet at one point after refraction. Rays PA, PB which are in one tangential (or meridianal) section meet at T. Rays PA, PC which are in one sagittal section meet at S'. As we shift the tangential section from AB towards CD, we find T shifting to T''. Thus TT' is a line image. Similarly SS' is another line image. TT' and SS' are called the *tangential* and *sagittal images* of P ; their lengths are perpendicular to the tangential and sagittal planes respectively. The error is called astigmatism and the images are called astigmatic images of P. Astigmatism is present in small aperture lenses and mirrors also if *inclination of the chief ray* from the principal axis is large.

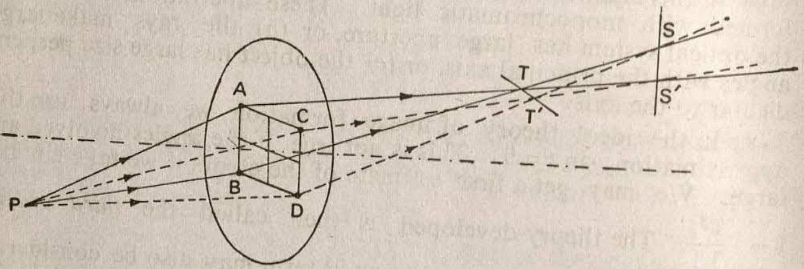


Fig. 5·2.

Showing astigmatic line images TT', SS' for object point P

(*iii*) **Coma.** In Fig. 5·3, let the lens be corrected to have zero spherical aberration, which means that for an axial point Q all zones of the lens form image at Q'. Now, consider a slightly *off-axis* object point P and a narrow outer zone of the lens shown by $ABCD\ A'B'C'D'A$.

---

*Largeness of aperture is to be considered relative to the focal length. Thus a lens of 1 cm diameter and $f=5$ cm has 'larger' aperture, than a lens of 5 cm diameter and $f=200$ cm.

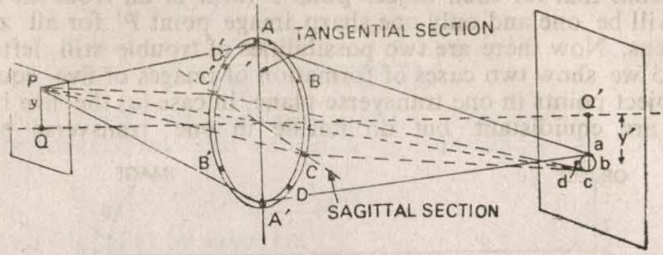
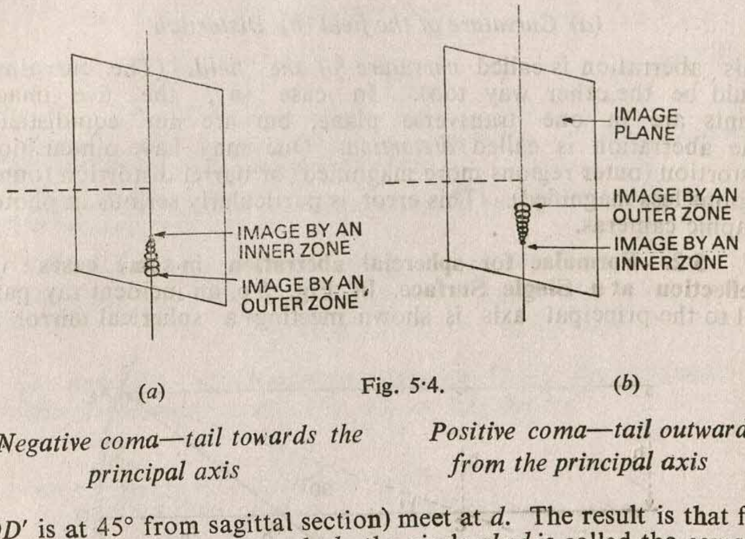


Fig. 5·3

Showing how rays through a given zone produce a comatic circle

The pair of rays *PA*, *PA'* (in tangential section) meet at *a*. The rays *PB*, *PB'* (*BB'* at 45° from tangential section) meet at *b*; the rays *PC*, *PC'* (sagittal section) meet at *c*, the rays *PD*, *PD'* (where *DD'* is at 45° from sagittal section) meet at *d*. The result is that for *P* we get a circular image *abcd*; the circle *abcd* is called the *comatic circle*. As the zone on the lens is taken more and more outwards, the comatic circle increases in radius. Not only this, the magnification $m = y'/y$ and hence height y' of the image (Fig. 5·3) also changes as we take different zones. These two facts combined give for the whole lens a coma-like image of *P* as shown in Fig. 5·4. This error is called *coma*, and is particularly serious in microscopes.



(*a*) Fig. 5·4. (*b*)

Negative coma—tail towards the principal axis

Positive coma—tail outwards from the principal axis

(*iv*) **Curvature of the field and** (*v*) **Distortion.** Let us consider a system free from spherical aberration, coma and astigmatism.

This means that for each object point P (near or far from the axis) there will be one and only one sharp image point P' for all zones ofthe lens. Now there are two possibilities of trouble still left. In Fig. 5·5 we show two cases of formation of images of five equidistant object points in one transverse plane. In case (*a*) the five image points are equidistant but do not lie in one transverse plane.

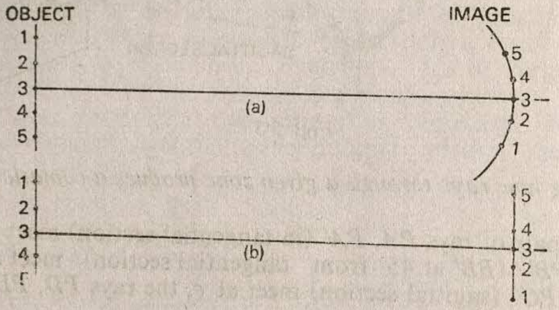


Fig. 5·5.

(*a*) *Curvature of the field* (*b*) *Distortion*

This aberration is called *curvature of the field*. (The curvature could be the other way too). In case (*b*), the five image points are in one transverse plane, but are not equidistant. The aberration is called *distortion.* One may have pin-cushion distortion (outer regions more magnified) or barrel distortion (outer regions less magnified). This error is particularly serious in photographic cameras.

5·2. Formulae for sphercial aberration in some cases : (*i*) **Reflection at a Single Surface.** In Fig. 5·6, an incident ray parallel to the principal axis is shown meeting a spherical mirror at

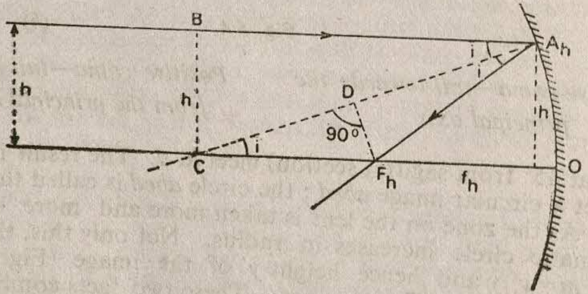


Fig. 5·6.

For deducing spherical aberration of a mirror

height h. The reflected ray intersects the axis at a point F_h. We call $OF_h=f_h$ the "focal length for rays in zone h". From Fig. 5·6, we readily see that

$$f_h = OC - CF_h$$

$$= R - \frac{CD}{\cos i} = R - \frac{R}{2\cos i}$$

If $h \to 0$, angle $i \to 0$, and then $f_h \to R/2$. This is the usual focal length for paraxial rays and we call it f_0. Substituting this for $R/2$ we get

$$f_h = f_0(2 - \sec i) \qquad \text{...(5·1)}$$

From Fig. (5·6) we note that

$$\sec i = \frac{CA_h}{BA_h} = \frac{R}{(R^2 - h^2)^{1/2}} \approx 1 + \frac{1}{2}\,\frac{h^2}{R^2}$$

$$= 1 + \frac{h^2}{8f_0^{\,2}},$$

that is the '*third order*' theory, neglecting *4th* and higher powers of h/R. Substituting in Eq. (5·1) and calling $f_0 - f_h$ the spherical aberration (SA) for object at infinity, we have

$$\boxed{SA = f_0 - f_h = +\frac{h^2}{8f_0}} \qquad \text{...(5·2)}$$

It may be more meaningful to see the *ratio* SA/fo. It comes out to be $(h^2/fo^2)/8$, and for $h/f_0 = 1/2$, it amounts to ~3%.

Example 5·1. Deduce the spherical aberration for a mirror of focal length 50 cm and aperture (*i*) 10°, (*ii*) 30°.

Solution. $\sec 10° = 1·017$, $\sec 30° = 1·155$

Using Eq. [5·1] $SA/f_0 = 0·017$, $SA/f_0 = ·155$

or $SA = 0·85$ cm $SA = 7·8$ cm. ⦿

(*ii*) **Refraction at a Spherical Surface.** Using Fig. 5·7 and a procedure similar to that with Fig. 5·6 we may deduce

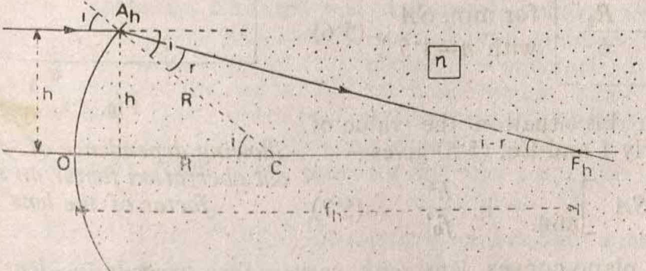


Fig. 5·7.

For deducing spherical aberration of a refracting surface

$$f_h' = OC + CF_h = R\left\{1 + \frac{\sin r}{\sin(i-r)}\right\} \quad \ldots(5\cdot3)$$

whence a 'third order' formula for SA (neglecting terms of fourth and higher order in h/R) comes to

$$SA = \frac{h^2}{2n(n-1)f_0'} = \frac{h^2}{2n^2 R} \quad \ldots(5\cdot4)$$

Example 5·2. For refraction at a spherical surface of $n=1\cdot5$, $R=20$ cm, evaluate the aberration for $h=5$ cm and $h=7$ cm.

Solution. From Eq. (5·4) the computation is direct. For $h=5$ cm.

$$SA = \frac{5^2}{2\times1\cdot5\times1\cdot5\times20} = 0\cdot28 \text{ cm.}$$

For $h=7$ cm, then

$$SA = 0\cdot28\times\frac{7^2}{5^2} = 0\cdot55 \text{ cm.}$$

⊙

(*iii*) **Refraction through a Thin Lens.** In the *third order theory* the spherical aberration of a spherical lens for objects at infinity is given by

$$SA = +\frac{h^2}{f_0'}\,[\phi\,(n, \beta)] \quad \ldots(5\cdot5)$$

where ϕ is a function of the refractive index n and the *shape factor* β defined by $\beta = R_1/R_2$, that is, the ratio of the radius of curvature of the *first* to that of the *second* face.

Fig. 5·8 shows a plot of ϕ against β for $n=1\cdot5$. It shows that there is a minimum value of ϕ for $\beta=-1/6$, which means—

$$\boxed{R_1 = -\frac{R_2}{6}} \quad \text{for min. } SA \text{ with } n=1\cdot5 \quad \ldots(5\cdot6)$$

For this situation the value of ϕ is nearly 1 and Eq. (5·5) gives—

$$\Big[\,SA\,\Big]_{\text{min.}} = +\frac{h^2}{f_0'} \quad \ldots(5\cdot7)$$

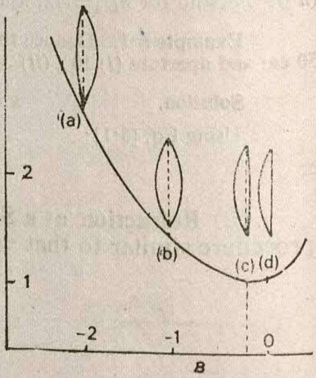


Fig. 5·8.

Showing dependence of spherical aberration factor on shape factor of the lens

A plano-convex lens *with convex face towards incident light* corresponds to $\beta=0$; Fig. 5·8 shows that its aberration is close to the minimum. For the same lens, with plane face towards incident light, β is $-\infty$, so that $\phi\,(n, \beta)$ and hence SA becomes quite large.

5·3. A General Principle for Reducing Spherical Aberration. There is a general principle that spherical aberration is minimum when the total deviation produced by an optical system is equally divided between various refractions that take place.

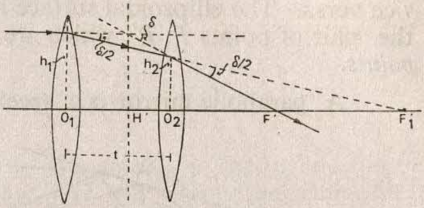


Fig. 5·9.

Equal distribution of deviations minimises SA

Applying this general principle to a combination of two lenses, in Fig. 5·9 the deviation δ is divided equally between the two lenses. Since deviation of a given ray by a lens is given by h/f' (see § 3·5), equal deviations mean

$$\frac{h_1}{h_2}=\frac{f_1'}{f_2'}$$

Also we notice from similar $\triangle^s$ that

$$\frac{h_1}{h_2}=\frac{O_1F_1'}{O_2F_1'}=\frac{f_1'}{f_1'-t}$$

Comparing the two values of h_1/h_2 gives

$$t=f_1'-f_2' \qquad \text{...(5·8)}$$

Thus, a combination of two lenses gives minimum spherical aberration if their separation is equal to the difference of focal lengths between the first and the second lens. Note that the result holds for incident rays parallel to the principal axis. But the chief ray in most cases so nearly satisfies this condition that rule (5·8) is generally applicable. Huygens' eyepiece (§ 6·4) satisfies this rule.

5·4. Aplanatic Surfaces and Aplanatic Points. Consider a setting in which *a very wide cone* of rays from a point P comes to sharp focus at another point P'. The point concerned are called aplantic points and the surface or lens concerned is called aplantic surface or lens. For a given lens or reflector this condition of zero spherical aberration for a wide cone of rays applies *for a particular pair of conjugate points* only, not for all the conjugate pairs of points.

In Fig. 5·10 let P and P' be a pair of points and consider the locus of a point A such that

$$PA+AP'=\text{Constant}=K \qquad \text{...(5·9)}$$

This locus of A is an ellipse. One of the properties of the ellipse is that PA and AP' make equal angles with the normal at A.

Therefore, if the ellipse is rotated about the axis PP' and the surface so obtained (ellipsoid) is silvered, we get a mirror such that *all* the rays from P, whatever their angle, converge sharply at P' and vice versa. The ellipsoidal surface is called an *aplanatic mirror*, and the pair of points P, P', which are its focii, are called the *aplanatic points*.

A parabolic mirror is a special case of the above one, where one aplanatic point is its focus, the other lies at infinity. Such mirrors are widely used in motor headlights, lighthouses, etc.

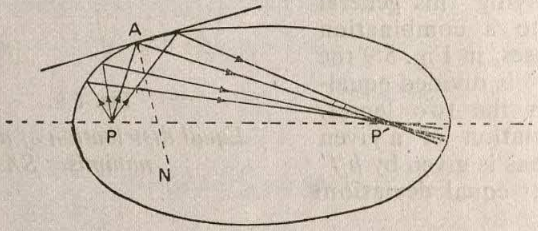


Fig. 5·10

The ellipsoid of revolution as an aplanatic reflector

In Fig. 5·11 we show a spherical medium with centre O, radius r and refractive index n, surrounded by air. Consider an object point P at distance $OP=r/n$. Consider *any* ray PA from P. It is refracted along AB. Let BA produced meet the line OP at P'. We propose to calculate OP'.

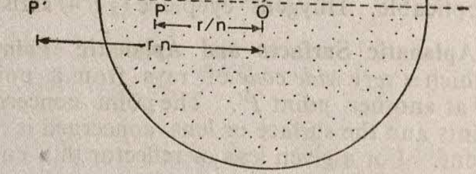


Fig. 5·11.

Aplanatic points of a sphere

In △ OPA,

$$\frac{\sin\theta}{\sin\alpha}=\frac{OA}{OP}=\frac{r}{r/n}=n.$$

But by Snell's law $\sin\beta/\sin\alpha=n$. Hence it follows that $\beta=\theta$.

Then in $\triangle\ AP'P$

$$\phi=\theta-(\beta-\alpha)=\beta-(\beta-\alpha)=\alpha$$

This gives

$$\frac{OP'}{OA}=\frac{\sin\beta}{\sin\phi}=\frac{\sin\beta}{\sin\alpha}=n \Rightarrow OP'=n.r \quad ...(5\cdot10)$$

We note that distance OP' is independent of the inclination (θ) of the incident ray *PA* from the axis *PO*. No approximation is involved. So all rays from *P*, *whatever their slope*, will, after refraction, appear to diverge from P'. These conjugate points P, P' are called *aplanatic points of the sphere.**

If we draw two spheres around *O* with radii *OP* and OP', then we shall find that for every object point on the first sphere there will be an image point on the second sphere free from spherical aberration. These two surfaces are called *aplanatic surfaces*. (They are not plane surfaces.) The linear magnification is $rn\div r/n=n^2$, and surface magnification is n^4.

Practical use of aplanatic points of a sphere is made through the use of (*i*) an oil immersion lens, and (*ii*) a 'meniscus' lens.

Oil Immersion Lens.** In this 'lens' a hemispherical lens of radius r and refractive index n is supported over an oil of the same index n and depth equal to r/n. The arrangement is shown in Fig. 5·12. As discussed above, for an object at P (distance r/n from centre of the sphere) all the rays in the wide cone appear to diverge from a single point Q (distant rn from centre of the sphere). There is thus no spherical aberration.

One can see that the semi-angle θ of the cone of incident rays from P entering the lens is given by $\tan\theta=n$, while the semi-angle θ' of the emergent cone of rays is given by $\tan\theta'=1/n$, so that the cone of rays converges by a factor $\tan\theta/\tan\theta'=n^2$.

---

*See § 1·6, Ex.1.6. A full discussion in the light of Fermat's principle is given there for this case.

**The advantage of oil-immersion of the object may also be seen in the light of Lagrange-Helmholtz law (Eq. 2·20). It increases the magnification. [See Example 2·4].

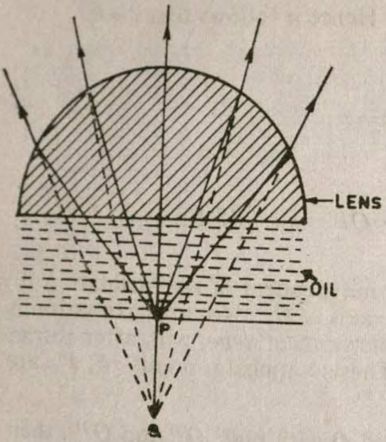


Fig. 5·12
Oil immersion objective as a practical use of aplanatic points of a sphere

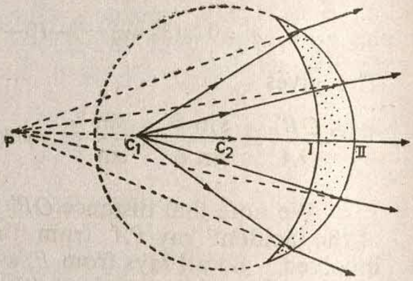


Fig. 5·13
A specially designed meniscus lens is aplanatic for a pair of points C_1, P'

Aplanatic Meniscus Lens. Fig. 5·13 shows this case. C_2 is the centre of curvature of the surface II of radius r. Choose C_1 such that distance $C_1C_2=r/n$ and choose surface I such that C_1 is its centre. Now if C_1 is the object point, all rays from it are normal to surface I and enter the lens without deviation. At surface II they bend so as to form a virtual image without spherical aberration at P', where $C_2P'=r.n$.

Often the meniscus lens is used together with an oil immersion lens in microscope objectives. You may turn Fig. 5·13 anticlockwise by a right angle and place it above Fig. 5.12, making C_1 coincide with Q [Fig. 6·12 (*a*) shows one such arrangement.]

Example 5·3. Calculate the radii of aplanatic surfaces for a refracting sphere of $n=1{\cdot}60$, $R=3{\cdot}2$ cm. Calculate also the magnification of the image in area.

Solution.

R (internal) $=3{\cdot}2\div1{\cdot}6=2{\cdot}0$ cm.

R (external)$=3{\cdot}2\times1{\cdot}6=5{\cdot}1$ cm.

Surface magnification $=n^4=6{\cdot}5$.

⊙

Example 5·4. An oil immersion objective and a meniscus lens are used in succession, each having $n=1{\cdot}60$. Calculate the semiangles of the incident and emergent cone of rays.

Solution.

For the incident cone, $\tan\theta=r\div(r/n)=n=1{\cdot}60$

For the intermediate cone, $\tan\theta'=r\div(rn)=1/n=0{\cdot}667$

In the meniscus lens θ' now becomes the incident cone. If θ'' is the emergent cone,

$$\tan\theta''=\frac{1}{n^2}\tan\theta'=0{\cdot}261$$

From tables we get

$$\theta=58^\circ\ ; \quad \theta'=33^\circ 42'\ ; \quad \theta''=14{\cdot}36'$$

[Note that the reduction of semi-angles of the cone reduces considerably the aberration by the later components in a microscope.]

5·5. Reduction or Elimination of Spherical Aberration. We now briefly record the different methods of reducing or eliminating spherical aberration.

(i) Reducing the Lens Aperture. Reducing the aperture h reduces SA considerably. But it also reduces intensity of the image and resolving power of the instrument (see Chapter 15).

(ii) "Bending" of a Lens. For given n and given f of a lens the geometrical quantity $\left(\frac{1}{R_1}-\frac{1}{R_2}\right)$ is fixed. But individual radii R_1 and R_2 can be chosen variously. Spherical aberration is a function of the shape factor $R_1/R_2=\beta$ [see § 5·2 (iii)]. We can therefore choose β in such a way that the aberration is minimum (see Fig. 5·8). This process of choosing β is called *"bending" of the lens*. When the bending exactly suits the minimum aberration condition, the lens is said to be a *"crossed lens."*

(iii) Use of a Separated Doublet. When a separated doublet is used, the separation between the lenses for minimum spherical aberration is given by Eq. (5·8) :

$$t=f'_1-f'_2$$

Additional advantage comes if each lens is itself "bent" to give minimum spherical aberration. This condition is used, for example, in Huygens' eyepiece (§ 6·4).

(iv) Contact doublet of convex and concave lenses of suitable shape factors. If K represents $[\phi(n,\beta)]$ of Eq. 5·5, so that $SA=Kh^2/f'$, then a contact doublet has zero SA when

$$\frac{K_2}{K_1}=-\left(\frac{f_2'}{f_1'}\right)^3 \qquad \ldots(5{\cdot}11)$$

(we omit the proof). This condition can be fulfilled together with the condition of achromatism because the shape factor is a free parameter for adjusting K values. But K depends also on u, v ; hence the condition holds for a given pair of conjugate points only.

(*v*) Use of Aplanatic Points. Property of aplanatic points is useful when the object can always be brought to a particular position, *e.g.*, in a microscope objective. In this case *spherical aberration is totally absent* ; but this holds only for one pair of conjugate points on the axis. For other pairs of conjugate points spherical aberration exists, but is small (coma is more serious here).

(*vi*) **Use of Specially Corrected Surfaces.** Mirror or lens surfaces may be specially "corrected" in such a way that rays from all zones come to a common focus. The surface is then not spherical and is called "aspherical". Such arrangements are used in all expensive telescopes and cameras using large aperture. However, the correction will be applicable for one pair of conjugate points only—hence for each special task special correction has to be worked out and executed.

An alternative is to use '*corrector plates*'. Fig. 5·14 shows the Schmidt corrector plate *P* used widely with large aperture reflector telescopes. As Eq. (5·2) shows, the rays falling on outer parts (larger *h*) come to focus earlier than those falling on the central part of the mirror. The Schmidt plate effectively compensates this by acting as a concave lens for rays at larger *h* and as a convex lens for rays at smaller *h*. Its positioning is at the centre of curvature of the reflector *R* to ensure that the effect is independent of the inclination of the chief ray.

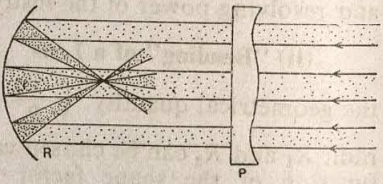


Fig. 5·14.

The Schmidt corrector plate P to correct spherical aberration of the reflector R

5·6. Elimination of Coma. Abbe's Sine Condition. In Fig. 5·15, for an axial object point *P*, the rays from lens zones defined by θ_1 and θ_2 form images at P_1' and P_2' respectively. (The fact that P_1' and P_2' do not coincide means that spherical aberration is present). For an off-axis object point *Q*, each lens zone would *normally* form a ring image, as shown in Fig. 5·3. But by a special design we may ensure a point image for each zone. In Fig. 5·15 Q_1' and Q_2' are the point images of *Q* for lens zones θ_1 and θ_2 respectively.

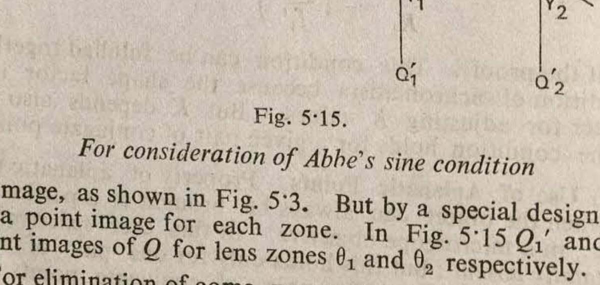


Fig. 5·15.

For consideration of Abbe's sine condition

For elimination of coma, we want the magnification y'/y to be independent of θ (even though the image plane may vary with θ).

Abbe showed that in any lens design which gives a true point image for off-axis object points for all zones θ, the relation

$$ny \sin \theta = n'y' \sin \theta' \qquad \text{...(5·12)}$$

must hold for all θ. This is called the *Abbe sine relation*. Absence of coma requires then that

$$\frac{\sin \theta}{\sin \theta'} = \text{constant for all } \theta \qquad \text{...(5·13)}$$

This is the Abbe sine condition for absence of coma. To check whether this holds for a given system one has only to do ray-tracing for an axial point source.

It may be noted that a design satisfying Eq. (5·13) for a given position P will not, in general, hold for other positions. Also the design may not be the one that gives minimum spherical aberration.*

The proof of Abbe's sine condition is rather tedious and we shall not give it. But Abbe's condition is certainly not derivable from the Lagrange-Helmholtz law (Eq. 2·20). Eq. (5·12) holds for all θ, but only in cases where point-for-pion image is ensured by the lens design for each zone. In contrast, the Lagrange-Helmholtz law is for spherical refracting surfaces, but for paraxial rays only. It says nothing for large θ values.

In an eyepiece, where $\theta \ll \theta'$, we may treat θ as proportional to the height h where the incident ray meets the principal plane H. Since the conjugate ray meets the plane H' at the same height h, the condition (5·13) reduces to

$$\frac{h}{\sin \theta'} = \text{constant} \qquad \text{...(5·14)}$$

It means that the principal 'plane' H' must be spherical with the image point as the centre. For objects at infinity (as in a telescope), the image point is F'; hence the H' is a sphere of radius f'. That makes the 'marginal' focal length equal to the paraxial one; so spherical aberration is also absent.

5·7. Elimination of Astigmatism. (*a*) *Use of stops*. With a single lens or mirror, astigmatism is avoided by using suitable stops to cut off rays making large slope angles with the axis. In Fig. 5·16 we show a concave mirror with C as its centre of curvature. For an object point P, beam PA_2 with its chief ray passing through C is a paraxial beam and the image for it will be free from astigmatism. If beam PA_1 is to form the image, its chief ray makes a large angle θ with the normal CA_1 at A_1. Image due to this will be astigmatic. A "*stop*" at C with a small central hole thus eliminates astigmatism.**

In the case of lenses also "stop" or stops introduced carefully minimise astigmatism. In Fig. 5·17, a stop at S with a central hole permits rays of large slope in object space, but small slope in image space like ray 7. A stop at S' does just the opposite. In this case, where $v \gg u$, the slope in the latter case (ray 2) is smaller, and hence S' is a better position for the stop. For each case a suitable position for the stop can thus be chosen.

---

* The aplanatic lens (§ 5·4), which has zero *SA* gives fairly low coma also; hence this system is widely used in microscope objectives.

**The slope is not to be measured from *CO* but from the normal at the point of incidence of the chief ray of a beam.

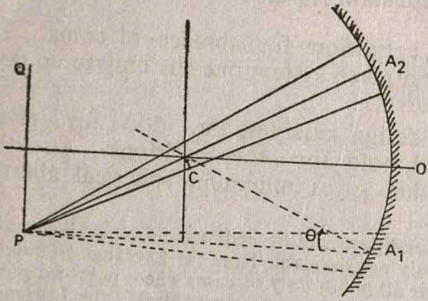


Fig. 5·16.

Choosing the proper position of a stop

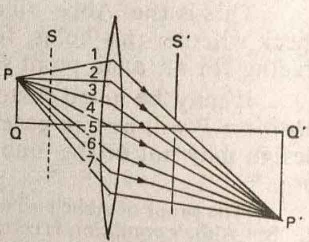


Fig. 5·17.

Use of a stop at the proper place eliminates coma

(*b*) *Petzval combination of lenses.* In Fig. 5·2 the axial separatoin of T and S (called the astigmatic difference) increases as inclination θ of the chief ray increases. Now, the astigmatic differences of a concave lens are opposite to those of a convex one. So by a suitable combination (see Fig. 5·18) one can arrange that S_1T_1 and S_2T_2 are equal in magnitude at all θ, though opposite in sign. Thus the resultant S and T surface coincide. This common image surfaces P is called the *Petzval Surface* after its discoverer. In good photographic cameras of large field angle Petzval combinations are used.

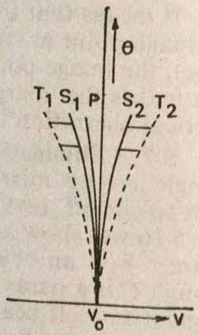


Fig. 5·18.

Opposite astigmatic errors of two components and the Petzval surface P

5·8. Elimination of Curvature of the field. (*a*) *Use of stops.* In a single lens, curvature is eliminated by giving the lens a meniscus shape and using a front stop (Fig. 5·19) somewhere near the centre of curvature of the meniscus shape (dotted curve). The net result is that as P goes away from the axis, we use outer parts of the lens and its meniscus shape brings these *outer* parts of the lens *nearer* the object, and hence their image points are pushed *farther* away. This is exactly what is desired to eliminate curvature of the field in a convergent system. Proper design will eliminate the field curvature error over a wide angle ; but the outer parts show marked astigmatism, since the rays emerge very obliquely from the second face, as can be seen in Fig. 5·19.

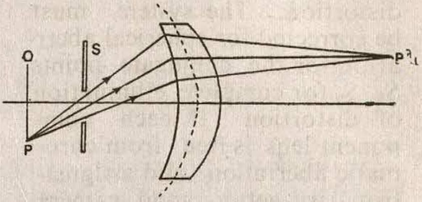


Fig. 5·19.

A stop and a meniscus lens for reducing curvature of the field

(*b*) *Petzval combination.* For a combination of two *equicurved* lenses, whether in contact or separated, Petzval has shown that curvature of field is zero when

$$n_1 f_1 + n_2 f_2 = 0 \qquad \ldots(5{\cdot}15)$$

where f_1 and f_2 are focal lengths of two lenses and n_1 and n_2 the refractive indices of their materials. Since refractive indices are always positive, it follows that f_1 and f_2 must have opposite signs. We get

$$f_1/f_2 = -n_2/n_1 \qquad \ldots(5{\cdot}16)$$

We recall that *for achromatism*, in a contact doublet, the condition is $f_1/f_2 = -\omega_1/\omega_2$(Eq. 4·12). In general, materials of higher n have also higher dispersive power ω. Hence elimination of curvature of the field and of chromatic aberration cannot be achieved simultaneously in a contact doublet.

Petzval condition (5·15) can be satisfied for the same material ($n_1 = n_2$) by choosing $f_2 = -f_1$. Placing two such lenses at a distance less than f_1 gives a convergent combination free from curvature of the field. Astigmatism will also be quite low since astigmatic differences are opposite for the two lenses, although not equal.

5·9. Reduction or Elimination of Distortion. Without a detailed analysis, we only state some results :

(*a*) A single thin lens shows no distortion.

(*b*) In case of doublets, a symmetrical doublet, with centre stop (Fig. 5·20) reduces distortion. The system must be corrected for spherical aberration for the conjugate points S_1, S_2 for complete elimination of distortion. If each component lens is free from chromatic aberration and astigmatism, we get a good camera lens or projection lens.

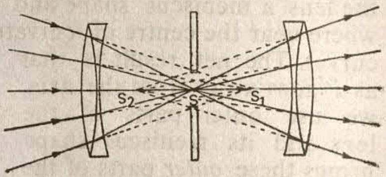


Fig. 5:20.
Symmetrical doublet and a centre stop eliminate distortion

PROBLEMS

5·1. Calculate the spherical aberration of a spherical reflector of $f=100$ cm and aperture diameter 60 cm.

5·2. Calculate the spherical aberration ratio $\triangle f/f$ for a spherical mirror whose aperture diameter is 1·4 times the paraxial f.

5·3. The spherical aberration of a lens is given by

$$\triangle f_h' = A.h^2/f_0'$$

where A is constant. Compare the aberration in the following three cases : (*i*) when central zone $h=0$ to 10 mm is used, (*ii*) when a peripheral zone $h=20$ mm to 24 mm is used, (*iii*) when the whole lens $h=0$ to 24 mm is used. Calculate also the relative amounts of light collected by the lens in the three cases.

5·4. Define 'shape factor' of a lens. Explain what is meant by 'bending' of a lens. A plano-convex lens is used to focus a distant object. Discuss the sharing of deviations δ_1 and δ_2 by the two faces when (*i*) the flat face is towards the object, (*ii*) the curved face is towards the object. Comment on the results in relation to 'shape factor' and 'bending'.

5·5. From the conditions for 'no chromatic aberration' and 'minimum spherical aberration' of a combination of two separated thin lenses, design a combination of equivalent focal length 5·0 cm.

5·6. Deduce the third order theory expression for the spherical aberration of a single refracting spherical surface [Eq. (5·4)].

5·7. An oil immersion objective uses a hemispherical lens of $n=1{\cdot}65$ and radius 0·80 cm. Deduce the position where the object should be located, and where its image would be formed. Deduce also the areal magnification of the image.

5·8. A thin mensicus lens is to be used as an aplanatic lens. If its larger radius of curvature is 20 cm and $n=1{\cdot}50$, deduce the radius of curvature of the other surface, and positions of the aplantic points.

5·9. A coaxial lens system is examined by ray tracing from a given axial point. The θ' values for different θ come out as follows :

$\theta=5°$	10°	15°	20°	30°	40°	50°	60°
$-\theta'=1°0'$	1°59′	2°58′	3°55′	5°44′	7°23′	8°49′	9°58′

Examine the system for comatic error. If the refractive indices of the initial and final media are 1·75 and 1·00 respectively, deduce the magnification. Can we conclude from the data that spherical aberration is absent ?

5·10. State Abbe's sine condition. Show that for object point at infinity it reduces to $h/\sin\theta'$=constant, where θ' is the angle of emergent ray corresponding to the ray incident at height h. If the system is to have zero *SA* too, then the principle "plane" H' must be spherical and of radius f'. Explain.

5·11. A concave mirror of radius of curvature 20·0 cm and large aperture is used with a 'stop' placed at the centre of curvature and having a small central hole. An object 8·0 cm in height is placed 15·0 cm away from the centre of curvature. Discuss how the stop affects (*i*) spherical aberration, (*ii*) astigmatism. Deduce '*u*' for the axial point and edge point of the object, compute the '*v*' values and draw a diagram to show the nature and extent of curvature of the field.

5·12. Define tangential section and sagittal section in relation to astigmatism, and explain how two sharp line images of a point object are formed in two planes. How may astigmatism be corrected for.

5·13. Instead of laws of reflection or refraction, use Fermat's principle (Chapter 1) to deduce Eqs. (5·1) and (5·3).

5·14. Consider the cases of (*i*) a telescope objective, (*ii*) a microscope objective, (*iii*) an eyepiece and (*iv*) a camera lens system. Which of the aberrations are most serious in each of these cases ? How are they generally minimised ?

5·15. Explain the principle behind the Petzval combination of lenses. Is the Petzval surface always a plane ?

6

Optical Instruments

The reader is familiar with the *image forming action of instruments* like the eye, camera, microscope, telescope, epidiascope, etc. What follows will deal with two related questions :

(*i*) What factors decide the apertures of the various components, the position of the eye, the total field of view, etc. ?

(*ii*) Apart from the aberrations discussed in the last two chapters, what other factors are to be considered in designing a viewing instrument so that it may show best detail ?

In effect most of our effort will be spent on the first question. The second question will be barely touched, because it would lead to more serious questions linked with the very nature of light, to be covered in Chapter 15.

6·1. Entrance Pupil and Exit Pupil. In Fig. 6·1, D' is a diaphragm with opening $A'B'$ in it and L is a lens in front of it. By

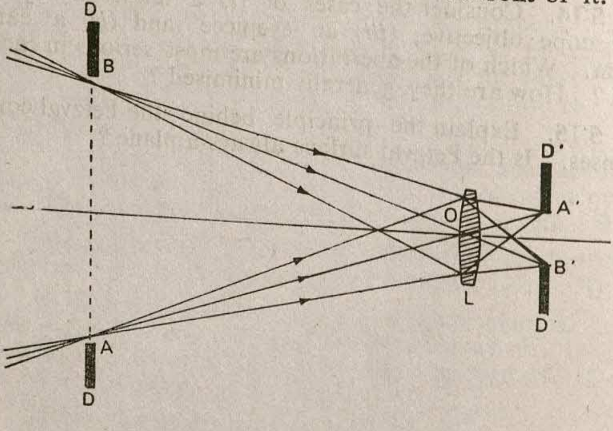


Fig. 6·1

Entrance pupil and exit pupil in a simple system backward drawing we obtain the points *A*, *B* in object space conjugate to *A′B′*.

If an eye or any instrument is placed to the right of *D′*, the width of the exit beam is limited by *A′B′*, which is called the *exit pupil* for this highly simple single lens viewer. The width of the entrance beam is *AB* which is called the *entrance pupil*.

If, however, you introduce a diaphragm in plane *D*, and its opening is *smaller* than *AB*, the situation changes. Imagine that the opening is (1/3) *AB*. Then the *entrance pupil* is (1/3) *AB* and that limits the exit pupil to (1/3) *A′B′*.

In this discussion it does not matter for us what is placed at the plane *D* or at plane *D′*— it may be a simple opening, or a convex lens, or a concave lens.* All that we assume is that *D* is the *first* element on the entrance side, and *D′* is the *last* element on the exit side.

In the intervening space between *D* and *D′* there may be any number of elements—convex lenses, concave lenses, or both. All that we assume is that *A′B′* is the image of *AB* by all the *intervening* elements. In other words, the *exit pupil* is the image of the *entrance pupil*.

What we place in the opening *AB* is certainly of interest to us. If it is a simple opening we see objects in the projection of cone *AOB*. If there is a convex lens there, we see object in a smaller cone. If there is a plane mirror there, we see objects in a cone behind ourselves.

It may be noted that an eye placed at *A′* does not *see* the images of *A*, it only *receives the light* which passes through position *A*. If the eye extends over range *A′B′* it *receives the light* which enters the instrument over the range *AB*. What the eye really *sees* are objects to the left of *AB*, and in this business the elements placed in the aperture *AB* and *A′B′* do play a serious part.

In most situations the eye itself is the exit pupil. A diaphragm with an aperture is provided in the instrument in the plane where the image of the front lens (not of the object) is made by the intervening lenses. This aperture is slightly larger than the aperture of the human eye ($\sim$2mm) ; therefore, the human eye sets the limit at this exit end.

An example will be useful. In a telescope the ratio of exit pupil to entrance pupil is equal to $1/m$, where m is the magnification. If a 40 cm diameter objective is used and magnification is 100, the exit pupil to use full objective aperture should be 4 mm. The eye would provide only about half the required pupil. Thus there is waste of the outer half of the objective. For full use of the 40 cm. diameter of the objective the magnification should be 200. For a camera or an epidiascope no such limitation exists.

6·2. Field of View. In Fig. 6·2, *E* is the eye and *AB* an aperture. In the absence of any aid (like the concave lens shown), the eye can view objects in the angular range *PEQ*. This is called the *field of view*. If the concave lens is added in plane *AB*, light

---

*Note that image of *AB* by a lens at the same place is *AB* itself. Similarly the image of *A′B′* by a lens at the same place is *A′B′* itself.

from a wider field $P'Q'$ can reach the eye. We say the *field of view is enlarged.* A convergent lens placed in plane AB would narrow the field of view.

In a single element like this the field of view is easily seen to be equal to the angle subtended by this lens aperture at the eye divided by the *magnification m* provided by the lens.

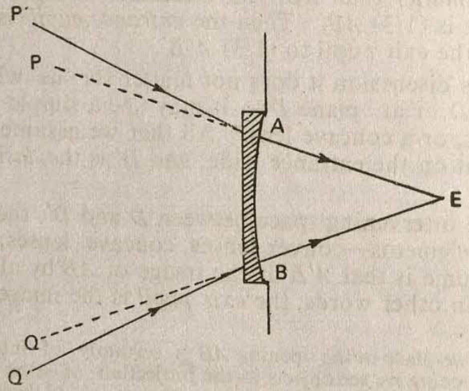


Fig. 6·2.

Field of view is enlarged by a concave lens

For a compound system the situation is slightly different Figs. 6·3 and 6·4 show the cases of a telescope and a microscope respectively. In each case the image of O by the eye-lens L_2 falls at E, the centre of the eye-ring (the exit pupil). If L_2 subtends angle θ_2 at the eye-ring, the field of view in Fig. 6·3 is given by

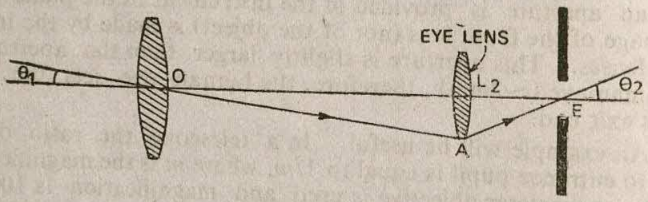


Fig. 6·3.

Field of view θ_1 is given by 1/m times angle AEL_2

$$\theta_1 = \frac{\theta_2}{m} \qquad \ldots(6{\cdot}1a)$$

where m is magnifying power of the telescope. Note that the aperture of objective O (entrance pupil) plays no role in determining the field of view. In a microscope (Fig. 6·4) the angle θ_2 is related to height y (rather than θ_1) and the magnification is defined by the ratio of θ_2 to y/D, where D is the least distance of distinct vision. We thus have

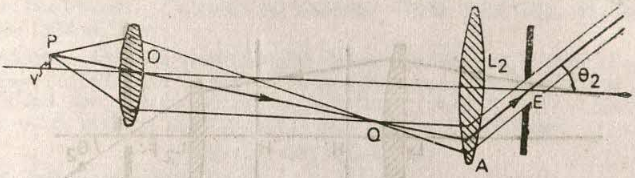


Fig. 6·4.

Field of view (y cm radius) in a microscope is given by $(\theta_2/m)\times D$

$$m=\frac{\theta_2}{y/D} \quad \text{or} \quad y=\frac{\theta_2}{m}\cdot D \qquad \ldots(6{\cdot}1b)$$

The field of view y is again dependent on θ_2, the angle subtended by the eye-lens at the eye-ring. As a near approximation, the distance L_2E is close to the focal length of the eye-lens, so that θ_2 is the ratio of the radial aperture to focal length for L_2. For a single lens this cannot be made large without serious aberrations and therefore considerable attention is paid to constructing eyepieces with multiple lenses so as to produce large field of view without much aberration.

6·3. Need for a Multiple Lens Eyepiece. For a given magnification m, the field of view is proportional to the angle θ_2 subtended by the eyelens at the exit pupil [Eq. 6·1]. For a single eye-lens to achieve high θ_2, there will be considerable aberrations. By using multiple lens eyepieces we therefore achieve two purposes.

(*a*) *Enlarging the field of view* : The field lens of the eyepiece effectively pushes the eyelens closer to the exit pupil, so as to enlarge the angle subtended.

(*b*) *Enabling us to minimise aberrations* : Firstly, the deviations of rays are divided over a large number of surfaces, so that spherical aberration is minimised. Secondly, for a separated doublet chromatic aberration can be reduced to zero or considerably minimised without involving different materials for the lenses.

In Fig. 6·5 we show the general function of an eyepiece. In the two-lens eyepieces the one nearer the eye is called *eye lens*, the other is called *field lens**

---

* The name itself suggests the role of the lens in enlarging the field.

To appreciate the functions (*a*) and (*b*) cited above, let us take up the case of Ramsden's eyepiece (§ 6·5). If $f_1'=f_2'=4·5$ cm, the focal length of the combination comes to 3·4 cm and distance of F' from the eyelens is 1·1 cm. This means that *for a given aperture* of eyelens the angle θ_2 obtained in the two-lens eyepiece is 3 times that obtainable by using a single lens of equivalent focal length (3·4 cm) and same aperture. Thus the field of view is considerably enlarged.

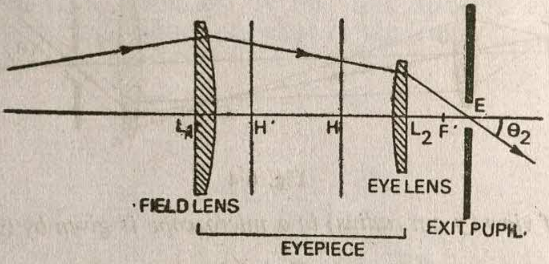


Fig. 9·5.

Explaining the function of a typical eyepiece (Ramsden's)

Also note that (for the same aperture) a single lens of $f'=3·4$ cm will have more aberration than the eyelens of $f_2'=4·5$ cm used in the two-lens eyepiece. The use of two separated lenses also reduces the chromatic and spherical aberrations as compared with a single lens.

Example 6·1. In a telescope with a single lens eyepiece of $f=6$ cm the objective is 106 cm from the eyepiece. Deduce the position of the exit pupil' magnification of the telescope, and size of entrance pupil if exit pupil is taken as 2 mm diameter. If diameter of the eye lens is 3 cm, deduce the semiangle of the field of view.

Solution. Exit pupil is at the image of objective by the eyelens (Fig. 6·1). Given $u=-160$ cm, $f=6$ cm, we get

$$v=\frac{106\times 6}{100}\text{ cm}=6·4\text{ cm}$$

For a telescope, magnification=angle subtended by eyelens at exit pupil ÷angle subtended by eyelens at the objective (Fig 6·3). Hence

$$m=106\div 6·4=\mathbf{16·7}$$

Size of entrance pupil÷size of exit pupil=magnification. Hence diameter of entrance pupil is

$$d(\text{entrance})=2\text{ mm}\times 16·7=\mathbf{3·3\text{ cm}}$$

From Fig. (6·3), semi-angle θ_1 is angle subtended by eyelens at the objective.

$$\theta_1=\frac{1·5}{106}=\mathbf{·0142\text{ radian}}$$

[Note that m obtained by considering angles is equal to the value f_0'/f_E' $=100/6$, obtained by considering the focal ratio.] ⊙

Example 6·2. In the case of Example 6·1 add a lens of $f=12$ cm. at the position of the primary image, giving it a 'suitable' aperture to use the full aperture of the eyelens. Calculate (*i*) aperture of this field lens, (*ii*) new semi. angle of the field of view.

Solution. Fig. 6·6 (not to scale) shows the scheme. With O as centre of the objective, OBE is the path of marginal rays when L_2 alone is present. With L_1 added, the path of marginal rays becomes $OABE'$. Here lens L_1 is the field lens added, making L_1+L_2 the two-lens eyepiece. We have

$$OL_1=100 \text{ cm}, \; L_1L_2=6 \text{ cm}$$

whence we deduce (check them) ;

$$L_1P=\mathbf{13\cdot6\ cm} \quad \Rightarrow \quad L_2P=7\cdot6 \text{ cm}$$

$$\text{aperture of } L_1=3\times\frac{13\cdot6}{7\cdot6}=\mathbf{5\cdot3\ cm}$$

$$\text{now } \theta_1=\tfrac{1}{2}\ \frac{5\cdot3}{100}=\mathbf{\cdot0265\ radian}$$

Thus the field of view increases to about 2-fold. You may note that lens L_1 here does not play a role in image formation (because its location coincides with the primary image) ; yet it enhances the field of view. The eye ring shifts from E to E', so that θ_2 also increases by the same factor as θ_1, keeping the magnification unchanged. ⊙

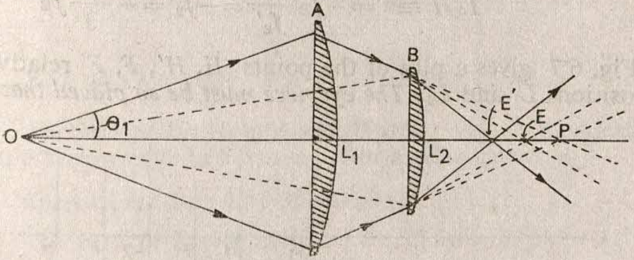


Fig. 6·6.

For Example 6·2. Introducing L_1 enhances the field of view.

6·4. Huygens' Eyepiece*.

(*a*) Description :

With f_1' and f_2' as focal lengths of the field lens and eye-lens, and t as their separation, the Huygens' eyepiece has

$$\frac{f_1'}{f_2'}=3,^{*} \quad t=\tfrac{1}{2}\,(f_1'+f_2')=2f_2'$$

Both components are plano-convex with plane faces of both towards the eye.

---

* The discussion under this head, particularly sub-sections (*f*) to (h), covers many important results of general applicability and should be studied with special care.

** Eyepieces with the ratio varying from 2 to 3 are used. Ratio 3 is preferred for use with telescopes and smaller ratios for use with microscopes. The separation is always kept $\tfrac{1}{2}(f_1'+f_2')$ to ensure achromatism.

(*b*) **Achromatism** : The condition is satisfied—

$$t\text{ (reqd.)}=(3f_2'+f_2')\tfrac{1}{2}=2f_2'\ ;\quad t\text{ (actual)}=2f_2'.$$

(*c*) **Condition for minimum SA** : The condition is satisfied—

$$t\text{ (reqd.)}=3f_2'-f_2'=2f_2'\ ;\qquad t\text{(actual)}=2f_2'$$

(*d*) **Actual f_1' and f_2' in terms of f_E'***—

$$f_E'=\frac{(3f_2')(f_2')}{3f_2'+f_2'-2f_2'}=\frac{3}{2}f_2'.$$

$$\therefore \qquad f_2'=\frac{2}{3}f_E'\ ;\quad f_1'=2f_E'.$$

The low focal length of the eyelens f_2' makes its contribution to spherical aberration quite large.

(*e*) **Positions of Cardinal Points** : From Eqs. (3·15) and (3·16),

$$L_1H'\equiv\alpha=f_E'\,\frac{t}{f_2'}=\frac{3}{2}\cdot 2f_2'=3f_2'=2f_E'$$

$$L_2H'\equiv\alpha'=-f_E'\,\frac{t}{f_2'}=-f_2'=-\frac{2}{3}f_E'$$

Fig. 6·7 gives a plot of the points H, H', F, F' relative to the lens positions L_1 and L_2. *The eyepiece must be so placed that plane F coincides with the primary image given by the objective of the instrument. Then the final image is at infinity and is seen with a relaxed eye.* Since F lies between L_1 and L_2, the Huygens' eyepiece is said to be negative.

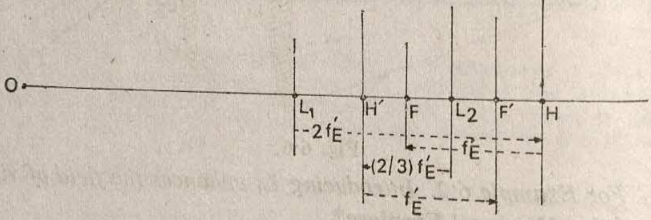


Fig. 6·7.

Cardinal points of Huygens' Eyepiece

(*f*) **Representative path of rays.** In Fig. 6·8 the cardinal points of the eyepiece are the same as in Fig. 6·7, but the lenses are explicitly shown. An incident beam converged by the objective is shown, tending to converge at P on the focal plane F. The two rays shown intersect the H plane at A and B ; these give the corresponding points A' and B' on plane H' (at equal heights). The image of O (centre of entrance pupil) by the eyepiece is formed at little beyond F', and that is where the eye-ring is placed (that is the centre of the exit-pupil).* We will leave the calculation, treating E as just a little beyond F', because $OH=u$ is $\gg f_E'$. Join $A'E$, and from B' draw a parallel line. That gives the emergent beam. Joining the intersections of the incident beam with L_1 and intersections of the emergent beam with L_2 gives the intermediate beam. (Notice that there is an intermediate convergence on plane H). Thus the full path of rays is obtained.

---

* The desired f_E' is of the order of 0·5 cm. The choice of f_1', f_2' is to meet that end.

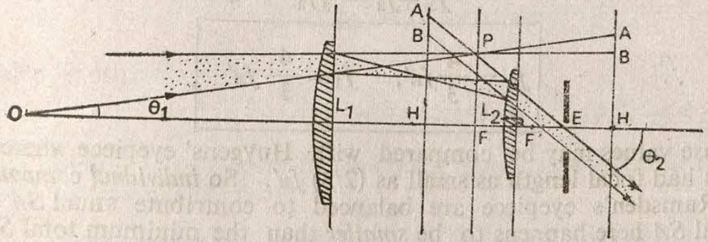


Fig. 6·8.

Path of a marginal beam in a Huygens' eyepiece

(*g*) **Apertures of the component lenses.** Fig. 6.8 defines the relative apertures of the two lenses. For a telescope case

$$\text{Aperture of } L_1=\theta_1\times OL_1\approx\theta_1\times f_0' \quad \text{...(6·5)}$$

$$\text{Aperture of } L_2=\theta_2\times L_2E\approx\theta_2\times L_2F'\approx\theta_2\times\tfrac{1}{3}f_E' \quad \text{...(6·6)}$$

Since $\theta_2/\theta_1=f_0'/f_E$ (=magnification), we get

$$\text{Aperture of } L_1\doteqdot 3\times\text{aperture of } L_2.^{**} \quad \text{...(6·7)}$$

(*h*) **Field of view.** Aberrations prevent using large lens apertures to increase θ_1. Comparing with a single lens eyepiece of same aperture as L_2, the two-lens system in effect makes θ_1 threefold. Thus the solid-angle field of view becomes 9-fold by use of the Huygens' eyepiece.

6·5. Ramsden's Eyepiece. This consists of two plano-convex lenses with the conditions.

$$f_1'=f_2', \qquad t=\tfrac{2}{3}f_2'$$

The curved surfaces of the lenses face one another as shown in Fig. 6·9. Now we examine the properties of this eyepiece.

---

* One may show that $F'E=f'_E(f'_E/f'_0)$ for a telescope case.

** Notice that the factor 3 came because $L_2F'=\frac{1}{3}f_E'$.

Condition for achromatism

t (required)$=\frac{1}{2}(f_1'+f_2')=f_2'$ from the data. But actual t is not this. Hence the eyepiece is not achromatic.

Condition for minimum ***SA***

t (required)$=f_1'-f_2'=0$ from the data. But actual t is not this. Hence the *combination* does not satisfy the condition of minimum *SA* for the given components.

The curved faces of both the lenses are put inwards to distribute the bending as evenly as possible over the four faces.

Actual f_1'**,** f_2 **in terms of** f_E'

$$f_E'=\frac{(f_2')(f_2')}{f_2'+f_2'-\frac{2}{3}f_2'}=\frac{3}{4}f_2'$$

$$\therefore \quad \boxed{f_2'=\frac{4}{3}f_E', \quad f_1=\frac{4}{3}f_E'}$$

Thuse values may be compared with Huygens' eyepiece where one lens had focal length as small as (2/3) f_E'. So *individual components* in Ramsden's eyepiece are balanced to contribute small *SA* ; the total *SA* here happens to be *smaller* than the minimum total *SA* in Huygens' eyepiece !

Position of Cardinal Points

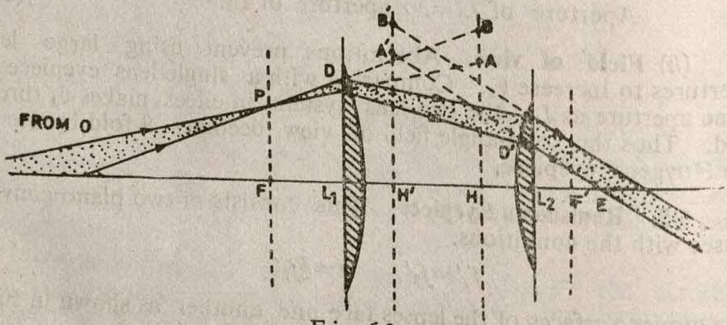
$$L_1H=\alpha=f_E'\frac{t}{f_2'}=\frac{1}{2}f_2'=\frac{2}{3}f_E'$$

$$L_2H' \text{ by symmetry}=-\frac{2}{3}f_E'$$

We will take a numerical case, with $f_E'=4{\cdot}5$ cm. That gives

$$f_1'=f_2'=6{\cdot}0 \text{ cm}, \ t=4{\cdot}0 \text{ cm}, \ \alpha=-\alpha'=3{\cdot}0 \text{ cm}.$$

Fig. 6·9 shows H at 3·0 to the right of L_1 and then F at 4·5 cm to the left of L_1. Thus the first focus F is outside the eyepiece.*



F ig. 6·9.
Cardinal points and path of a typical marginal pencil through a Ramsden's eyepiece.

---

[*] This is because t is kept below the value needed for achromatism. The sacrifice there comes as a benefit here,

For this reason Ramsden's eyepiece is called a *positive* eyepiece. One can use cross-hairs in plane F, and the eyepiece would magnify the primary image and the cross-hairs equally. With a graticule or a a micrometer drive for the cross-hair, the primary image can thus be measured.

Representative Path of Rays

Choose a point P on focal plane F. Let O be centre of the objective of the instrument (telescope or microscope). OP, when produced, meets the H plane at A. The conjugate of A is A', at the same height in plane H'. The next step is to locate E, centre of the eye-ring. Typically, let OF be 45 cm. That gives

$$u = P \text{ to } O = -49{\cdot}5 \text{ cm}$$

which leads to

$$v = H' \text{ to } E = \sim 5{\cdot}0 \text{ cm} \qquad (F'E = 0{\cdot}8 \text{ cm})$$

Joining A' to E gives the emergent ray corresponding to OP. The actual path is $OPDD'E$.

Another ray through P leads to locating B, hence B', and hence the emergent ray through B' parallel to $A'E$.

Apertures of the Component Lenses

If height $HA = h$, the aperture of the field lens is given by $h \times OL_1/OH = h$, since the distance $L_1H \gg OH$ in all cases.

The aperture of the eye-lens is given by $h \times L_2E/H'E$. In the example chosen, $H'E = 5{\cdot}0$ cm, and $L_2E = H'E - H'L_2 = 2{\cdot}0$ cm. That gives the aperture as $(2/5)\,h$. Thus the ratio of apertures is 5 : 2.

Field of View

Since distance L_1L_2 is much less than OL_1, we may consider the aperture ratio 5 : 2 as the enlargement of the field of view in angle (~6-fold in solid angle). In fact, the equivalent single lens would have focal length $f_E' = \frac{3}{4} f_2'$, and hence, for equal aberrations one would need to have its aperture even smaller than that of L_2 in the two-lens eyepiece. The field of view enlargement may thus be taken as nearer to $(5/2) \times (4/3)$, *i.e.*, over 3-fold.

6·6. Gaussian Eyepiece. This is the same as the Ramsden's eyepiece, except that a glass plate inclined at 45° to the axis is placed between the two lenses as shown in Fig. 6·10 and a hole is provided on the side to admit light. This enables us to illuminate the back side of the cross-hair. This eyepiece is used to assist mechanical setting of a spectrometer, and also when faint images in a dark background are to be *measured.*

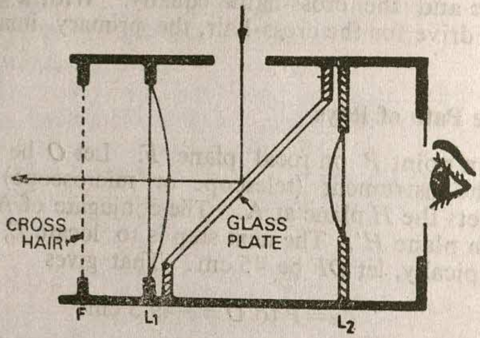


Fig. 6·10.

Gaussian Eyepiece

6·7. Kellner's Eyepiece. This eyepiece uses two plano-convex lenses with conditions

$$f_1'=f_2' \quad \text{and} \quad t=f_2',$$

thus only slightly differing from Ramsden's eyepiece where t was taken somewhat smaller ($2f_2'/3$). In this eyepiece both the curved faces are put towards the incident light (Fig. 6·11).

The eyepiece satisfies the condition of achromatism, but not the condition of minimum *SA*.

The relation of f_1', f_2' with f_E' is as follows :

$$f_E'=\frac{f_2'\times f_2'}{f_2'+f_2'-f_2'}=f_2'=f_1'$$

The components thus have focal lengths (3/4) times smaller than those for the equivalent Ramsden's eyepiece. Hence spherical aberration is more.

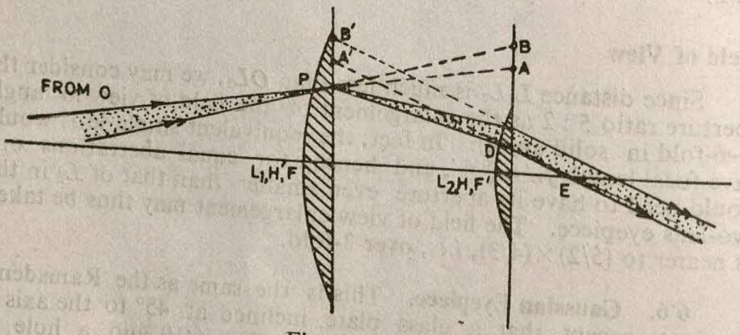


Fig. 6·11.

Kellner's Eyepiece has very simple location of its cardinal points

The cardinal points have positions given by

$$\alpha = f_E', \qquad \alpha' = -f_E' \qquad \text{(check)}$$

Thus H coincides with lens L_2 and H' coincides with lens L_1. When F, F' are plotted they coincide with L_1 and L_2 respectively (Fig. 6·11).

That F' coincides with L_2 has a special significance. Since the exit pupil E is just to the right of F', L_2E is very small compared with f_E'. This factor would appear to give almost 30 to 50 fold gain of field of view if considerations of §§6·4 and 6·5 are applied directly. But the exit pupil aperture is limited to ~3 mm diameter (due to eye pupil aperture ~2 mm dia.) and when L_2 is so close it cannot usefully have the aperture of the normal eye-lens. Hence in practice the gain in field of view is ~5 for use with telescopes, which may reach ~10 for use with microscopes.

Another characteristic of Kellner's eyepiece is that F coincides with lens L_1. The cross-hairs are therefore engraved on the lens face itself. In fact, in most cases a fine scale is engraved on L_1 for measurement. This advantage has one accompanying trouble—any dust particles depositing on L_1 are seen in sharp focus !

Example 6·3. A Huygens' eyepice having focal ratio 2·5 and satisfying achromatism is to be designed. Deduce focal lengths of the components and the cardinal points in terms of f_E'. In what respects does it differ from the Huygens' eyepiece with focal ratio 3 ?

Solution. $$f_E' = \frac{(2\cdot5 f_2')(f_2')}{2\cdot5 f_2' + f_2' - \frac{1}{2}(2\cdot5 f_2' + f_2')} = \frac{2\cdot5 f_2'}{1\cdot75} = \frac{10}{7} f_2'$$

$$\therefore \quad f_2' = \frac{7}{10} f_E', f_1' = \frac{7}{4} f_E', \qquad t = \frac{49}{40} f_E'.$$

Note that though *ratio* f_2'/f_1' is smaller than in standard Huygens' eyepeice, actual f_2' is slightly larger (hence smaller spherical aberration).

Next one may see that $L_1H = \alpha = \frac{7}{4} f_E', L_2H' = \alpha' = -\frac{7}{10} f_E'$

$$L_2F' = \alpha' + f_E' = \frac{3}{10} f_E' \; ; \; L_1F = \alpha - f_E' = \frac{3}{4} f_E'$$

The first focal plane F is still to the right of L_1. The eyepiece is thus negative. The quantity f_E'/L_2F', which determines enlargement of the field, is 10/3. This is large than in the normal Huygens' eyepiece ⊙

Example 6·4. Design an eyepice (f_1', f_2', t, eye-ring position, aperture ratio) to be used with a microscope of standard optical length 18·0 cm, given desired $f_E' = 4\cdot0$ cm, F at 0·5 cm in front of the field glass and F' at 0·8 cm behind the eye lens.

Solution. (Fig. 6·9 may be used for reference). Given

$$\frac{1}{f'_1} + \frac{1}{f'_2} - \frac{t}{f'_1 f'_2} = \frac{1}{4\cdot0}$$

$$-4\cdot0\left(1 - \frac{t}{f'_2}\right) = -0\cdot5$$

$$4{\cdot}0\left(1-\frac{t}{f'_1}\right)=0{\cdot}8$$

These lead to (check)

$$t=3{\cdot}9 \text{ cm}; \qquad f_1'=4{\cdot}9 \text{ cm}; \qquad f'_2=4{\cdot}5 \text{ cm}$$

Using Fig. 6·9. we now have E as the image of O, But

$$OF=18{\cdot}0 \text{ cm} \quad \Rightarrow \quad OH=-u=22{\cdot}0 \text{ cm}$$

That leads to

$$v=H'E=4{\cdot}9 \text{ cm} \quad \Rightarrow \quad F'E=0{\cdot}9 \text{ cm}; \qquad L_2E=1{\cdot}7 \text{ cm},$$

Now, aperture of field glass $=h \times OL_1/OH$

" " eye lens $=h \times L_2E/H'E$

$\therefore$ aperture ratio $r=\dfrac{OL_1}{OH} \times \dfrac{H'E}{L_2E}=\dfrac{15.0}{18{\cdot}0} \times \dfrac{4{\cdot}9}{1.7}$

or $r=2{\cdot}3$

Note: The design is still incomplete, because shape factors of the lenses are not specified. Also note that for use with a telescope the details of L_2E and r will change.

⊙

6·8. Comparison of Huygens' Ramsden's and Kellner's Eyepieces. In Table 6·1 we give a comparison of the components and performances of the three eyepieces. We take f_E' as the standard in terms of which the focal lengths of components and the separations are expressed. Some information not earlier discussed is also included in this table.

TABLE 6·1

Comparison of the Eyepieces

Property	Huygens' Eyepiece	Ramsden's Eyepiece	Kellner's Eyepiece
1. Focal lengths of components	$\frac{2}{3}f_E'$, $2f_E'$	$^4/_3f_U'$, $^4/_3f_E'$	f_E', f_E'
2. Separation	$^4/_3f_E'$	$\frac{2}{3}f_E'$	f_E'
3. Whether achromatic	yes	no	yes
4. Whether min. *SA*	yes	no	no
5. Location of F (hence cross-hairs)	inside	outside	on field lens
6. Field of view compared with single lens	∼3	∼3	∼5 to 10
7. Actual *SA* comparative	largest	least	medium
8. Coma and astigmatism	larger	least	?
9. Curvature of field in final image	Convex towards the eye	almost flat	?
10. Instruments where widely used	Telescopes and microscopes	Spectrometers, Telescopes	High power microscopes

6·9. Compound Microscope. While in a telescope the objective is a single lens (or a contact doublet) of large focal length, in a compound microscope the objective is a complex unit of very small focal length f_0'. Our attention therefore shifts to the objective lens-system.

The magnifying power of a microscope is given by

$$m=\frac{25.x'}{f_0'f_E'}\qquad (x', f_0', f_E' \text{ in cm})\qquad \ldots(6{\cdot}11)$$

where x' is the distance corresponding to OF in Fig. 6·8. Since the distance $f_0'+OF$ (called the *optical length of the microscope*) is standardised to 18 cm, there is little margin for increasing x'. Also, considerations of various aberrations and the fact that the exit pupil must not be smaller than pupil of the eye limit the lower practical value of f_E'. Therefore, all high power microscopes aim at securing lower value for f_0'.

High power microscope objectives require also a *large cone* of light to enter them, because, as we shall see later, the viewing of details ('resolution'—see Chapter 15) depends upon this. With large cones of light, spherical aberration and coma cause serious troubles. To avoid spherical aberration we use a large number of lenses, including aplanatic lenses (see § 5·4). To avoid chromatic aberrations we use achromatic lenses (see § 4·4) which are over-corrected to balance the chromatic errors of the first two aplanatic lenses also. Finally, to avoid coma we arrange the components so as to satisfy the Abbe sine condition (§ 5·6). The designing of high power microscope objectives is thus a highly complex affair. Not only this. The *use* of these objectives involves accurate setting of the object and special methods of illumination. Therefore, a 'condensing system' for light and a 'stage' for sample mounting are integral parts of all high power microscopes. Fig. 6·12 shows two typical microscope objectives used.

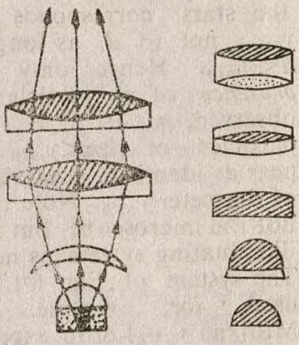


Fig. 6·12.

Typical high power microscope objectives

The smallest size of particles (or distance between two close points) which can be resolved (viewed separately) with a given microscope is given by

$$y_{min}=\frac{1{\cdot}22\,\lambda}{\text{Numerical Aperture}}\qquad \ldots((6.12)$$

where λ is the wavelength of light used and *numerical aertpure* is the product of the sine of the semi-angle of cone of light received by the objective and the refractive index of the medium in which the object is placed. Details are discussed in Chapter 15. But we may see that use of smaller λ would enable us to observe smaller details. That is the secret of the *electron microscope.*

6·10. Ultra-miscroscope. Let y_{min} be the smallest diameter of the object which can be resolved by a given microscope. Then if there are particles of diameter d *smaller* than y_{min} suspended in air or a liquid, this microscope is not able to see them ordinarily. Such particles are called *ultra-microscopic particals.*

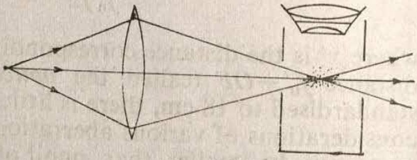


Fig. 6·13

One method of illumination in an ultramicroscope

If the medium is illuminated from a side (Fig. 6·13) so that no part of the *direct beam* enters the microscope objective, then the scattered rays from the particles only reach the microscope. The particles are then seen as bright stars in a *dark background.* This is the principle of the *ultra-microscope.* The apparent size of the stars corresponds to y_{min}, not to d, as long as $d<y_{min}$. Hence only the *presence* of a particle is observed, *not its size.* All particles of size $<y_{min}$ appear as identical sized discs of diameter y_{min}. Note that not the microscope but the illuminating system is new. The system of Fig. 6·13 is used, for example, in Millikan's oil-drop experiment. The condensing system of a regular ultra-microscope is shown in Fig. 6·14. The incident beam from the bottom enters through the annular opening in the diaphragm D, and the paraboloid condenser PP gives a hollow cone of light CC, such that all *direct* light goes outside the microscope objective. This is called *dark field illumination.* Only the scattered light from the object placed on plate S enters the microscope.

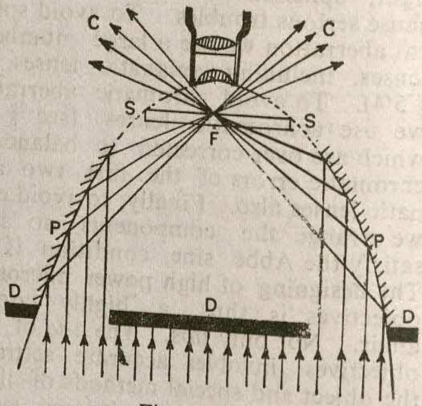


Fig. 6·14.

Another Condensing system in an ultramicrosope

6·11. Telescope. Telescopes are used to view very distnat objects, which subtend very small angle θ_1 at the eye (Fig. 6·3). The telescope makes image of such objects so as to subtend an angle θ_2 at the eye, where θ_2/θ_1 is called the 'magnification', m. If f_0' is the focal length of the objective and f_E' that of the eyepiece, the angular magnification is given by $m=f_0'/f_E'$, whatever the eyepiece used.

Since f_E' cannot be too small in view of aberrations, large magnifying power requires large f_0', *i.e.*, objectives of very large focal lengths. The most serious error involved in telescope objectives is the *chromatic aberration*. Hence refracting telescopes must have achromatic doublets as objectives (§ 4·4). All large telescopes are, in fact, of the reflecting type, where chromatic error is absent.

We further find that objectives of large focal length must also have large aperture. This can be seen from the Helmholtz-Lagrange law, Eq. (2·20). The *entrance pupil O* (aperture of the objective) and the *exit pupil E* (ultimately the eye) form a conjugate pair of planes. Hence, if ray from O at angle θ_1 emerges at E with angle θ_2, the transverse dimensions at O and at E must satisfy the relation

$$y_0 \tan \theta_1 = y_E \tan \theta_2$$

Now y_0/y_E is the appropriate ratio of diameters of the objective and the exit pupil (d_0/d_E), and $\tan \theta_2/\tan \theta_2 \approx \theta_1/\theta_1$ is the magnification m, which in turn equals f_0'/f_E'.

$$\therefore \qquad \frac{d_0}{d_E} = m = \frac{f_0'}{f_E}$$

Now, if the aperture d_0 of the objective is less than $m\,d_E$, then full capacity to see details is not utilised (empty magnification). Also, if d_0 is greater than $m\,d_E$ then full capacity of the objective to show details is not utilised (empty resolution). Hence for 'normal' conditions,

$$d_0 = m.d_E$$

Example 6·5. If aperture of the eye is taken as 2 mm and least distance of distinct vision is 25 cm, calculate the 'numerical aperture' of the unaided eye. The best numerical aperture obtained in a microscope approaches a value 2. How much should the normal magnification of such microscope be ?

Solution. Semiangle $=0{\cdot}1$ cm/25 cm$=0.004$ radian. For air $n=1$. Hence

N.A. of the eye=**0·004 radian**

Considering $1{\cdot}22\lambda$ term in Eq. (6·12) common between eye and a microscope, the detail resolved by the *objective* of the microscope is smaller by factor

$$\frac{\text{N,A. of microscope}}{\text{N.A. of the eye}} = \frac{2}{.004} = 500$$

So magnification should be >500 fold if the microscope is to use full resolution. 'Normal' magnification $=500$. ⊙

Example 6·6. Let the focal length of an eyepiece be 2·5 cm. For a telescope of magnification 100, deduce the focal length of the objective and useful diameter of the objective, given $d_{eye}=2$mm

Solution.

Given $m=\frac{f_0'}{f_E'}=100$ and $f_E'=2{\cdot}5$ cm

$\therefore$ $f_0'=250$ cm.

For optimum use, $\frac{d_0}{d_E}=m$; and $d_E=2$ mm (given)

$\therefore$ $d_0=20$ **cm.** ⊙

PROBLEMS

6·1 An aperture of diameter 6.0 cm is placed 20.0 cm in front of the eye. Calculate angular diameter of the field of view. If a concave lens of focal length 30.0 cm is placed on the aperture, deduce the new field of view.

6·2. Discuss the fields of view of (i) a plane mirror, (ii) a concave mirror, and (iii) a convex mirror.

6·3. Show that in a telescope the field of view depends on magnification *m* and the angle subtended by the eyelens at the eye, and is independent of aperture of the objective. What is then the need for having a large aperture of the objective ?

6·4. In a microscope, the first image is formed 30.0 cm behind the objective, and an eyepiece of $f_E'=5{\cdot}0$ cm is set for viewing with a relaxed eye. Deduce the position of the *eye-ring* relative to F' of the eyepiece.

6·5. Why is a double-lens eyepiece better than a single lens eyepiece of equivalent focal length ? Explain.

6·6. A telescope objective is rendered achromatic by using a contact doublet, but an eyepiece to be rendered achromatic uses separated lenses. Discuss.

6·7. Deduce the composition of a Ramsden's eyepiece of equivalent focal length 4.2 cm. Derive and plot its cardinal points. Show the path of a pencil of rays through it, covering the full aperture of the eye-lens.

6·8. Distinguish between a '*positive*' and a '*negative*' eyepiece. If a cross-wire is to be used with a Huygens eyepiece, show where it should be located. What are the limitations, if any, of such use ?

6·9. Deduce f_1', f_2', t, eye-ring position and aperture ratio of an eyepiece to be used with a telescope whose objective has focal length 60·0 cm, given $f_E'=4{\cdot}5$ cm, F at 1·0 cm in front of the field lens and F' at 0·5 cm behind the eye lens.

6·10. In Fig 6·8 a pencil of rays corresponding to an intermediate image point *P* near the edge of the field of view is traced.

Draw the corresponding diagram for a point at the centre of the field of view (*i.e.*, P coinciding with F).

9.11. Broadly explain the functions of the components of the high power microscope objective of Fig. 6·12 (a) which uses an oil immersion lens, a meniscus lens and two contact doublets.

9.12. Aperture of the eye is ~ 2 mm in diameter and effective λ for vision is $\sim 5 \times 10^{-5}$ cm. Deduce the smallest particle size which can be resolved by an appropriate microscope of magnifying power 60. (Take least distance of distinct vision as 25 cm). What is to be 'appropriate' about the microscope ?

7

Speed of Light

Galileo suggested the following method for finding the speed of light : Two men with lanterns stood over two hills, a measured distance apart. The first uncovered his lantern and started a clock, and the second uncovered his lantern on seeing the light from the first. When the first man saw the second man's light, he stopped the clock. The experiment failed. But it was not a complete failure because it showed that the speed of light was too great for measurement with short distances and crude timings.

These days *microwave pulses* are reflected from the moon (distance $\sim 8\times10^8$ m), and from the time of return of the pulse one may get the speed of light c. But other methods have given c so accurately that these reflections are now used to measure moon's distance and its variations.

Roemer in 1675 used a method not needing small time measurement. The satellites of planet Jupiter show eclipses once in every 7 days which is their time of revolution about Jupiter. But the period was found by Roemer to be 11 minutes longer than the mean when earth was farthest from Jupiter and 11 minutes shorter when earth was nearest. Roemer attributed this 22 minutes difference to the time taken by light in travelling the diameter of earth's orbit. This diameter being 3×10^{11}m, Roemer calculated the speed c :

$$c=\frac{3\times10^{11}m}{22\times60\ \text{sec}}=2.3\times10^{8}m/\text{sec}$$

Later measurement showed that the time difference was actually 16 min. 20 sec., and hence the value needed revision. But that is not the point. A great step had been taken in establishing that speed of light, though large, is not infinite.

Half a century later, in 1726, *Bradley* noted that the fixed stars showed apparent displacements (aberrations) in angular positions over the year, and he attributed it to the finite speed of light, obtaniing a value in agreement with that obtained by Roemer's theory. Both these may by listed as *astronomical methods*.

In fact the word 'method' is probably not appropriate. A natural observation was explained by an assumption in each case. The assumption may or may not be valid. For instance, one could at least imagine that the moons of Jupiter really did not go round Jupiter with constant period over earth's year for some reason. To say that the periods are not likely to vary and the differences in *our* observations are due to the extra distance $3 \times 10^{11} m$ travelled by light when we are farthest from Jupiter, is a theory. The best support for it was that Bradley, in interpreting another astronomical phenomenon, got a value of c in close agreement with Roemer's.

The first *direct* measurement of the speed of light from an experiment of sending a signal over a distance and measuring the time of travel came over a century after Bradley's work. That was in 1849. The experimenter was *Fizeau* using path lengths ~ 10 miles. Soon after that *Foucault* discovered a method which involved distances of only a few metres and which was therefore suitable for measuring speeds of light in material media. Another *terrestrial method* was devised by Michelson, around 1880-83 using long paths to get a high precision determination of c. An entirely different method was later discovered which used path lengths of only a few metres and yet gave accuracy of the same order as Michelson's. It is the *Kerr cell* method, devised in 1925 by *Karolus* and *Mittelstaedt* and improved, among others, by *Anderson*.

We shall now briefly describe the outlines of the methods of Fizeau, Foucault, Michelson and Anderson.

7·1. Fizeau's Method. The optical arrangement of Fizeau's method is shown in Fig. 7·1. The heart of the method is the toothed wheel W shown separately in the lower part of the figure. The teeth and the gaps in between made an alternating shutter when the wheel was made to rotate. In one of the experiments the wheel had $\sim$500 teeth and rotated at $\sim$10 rev/sec. Thus time intervals of 1/5000 sec were obtained.

Now, the image of a strong source A is formed at X by the lens L_1 and 45° inclined half-silvered glass plate P. X is at the rim of the toothed wheel W, described above. If there is a gap at X, light a passes through, is rendered parallel by lens L_2, and after travelling distance of *several kilometres* falls on the system L_3M. The focal point of lens L_3 lies on mirror M and the centre of curvature of M lies on L_3. The net result is that the beam is returned by the system on its own path and is collected by lens L_2 at X. If there is a gap in wheel W at that moment, it passes through. A part of this light is passed by the plate P, and is collected by lens L_4 at B. This image is seen through the eyepiece E.

All the four lenses are compound achromatic systems ; they are shown in the figure as single lenses for clarity only.

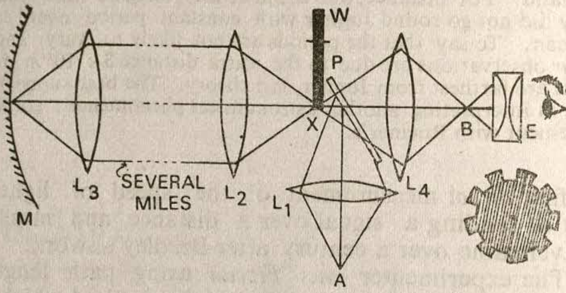


Fig 7·1.

Scheme of Fizeau's method. The toothed wheel design also shown separately

To start an experiment, a 'gap' in W is brought at X and optical setting is completed to get the return image of A at B. Then the wheel is rotated and its speed is gradually increased. If the time taken by light to travel distance $XM=d$ and back is equal to the time in which the wheel rotates to replace a 'gap' by a 'tooth' at X, then the image at B is cut off or 'eclipsed'. At this stage the rotation speed of wheel W is kept constant and the number of revolutions per second (m) is measured.

If the number of teeth in W is n, we have

$$c=\frac{\text{distance}}{\text{time}}=\frac{2d}{1/(2mn)}=4mnd \qquad \ldots(7{\cdot}7a)$$

If the rotation of W is made faster still, a stage comes where the image B becomes bright again. It can be at once seen that if this time the number of revolutions of W per second is m_1, we have

$$c=2m_1nd\,; \qquad \ldots(7{\cdot}7b)$$

and so on.

It is to be realised that eclipse or full brightness condition do not occur suddenly ; as the speed is gradually increased, the intensity goes on falling till a total eclipse occurs. As Example 7·1 below shows we cannot increase m or n by wide margins. Hence d has to be several kilometres.

Example 7·1. In one of his later experiments Fizeau used 720 teeth in his wheel, and 8 kilometer distance. The radius of the wheel may be taken as 20 cm. Calculate (i) the spacing between successive teeth, (ii) the frequency of revolution for the first eclipse, (iii) the centrifugal acceleration at the rim of the wheel.

Solution.

spacing$=2\pi r/720=0{\cdot}175$ cm.

frequency$=c/4nd=13$ rev/sec.

centrifugal accln.$=\omega^2 r=(2\pi\ 13)^8.\ 20$

$$=1{\cdot}33\times10^{5}\ \text{cm/sec}^{2}\approx140\ g$$

For the next eclipse the frequency will be 3 times and acceleration at the rim will be 9 times this value). ⊙

7·2. Foucault's Method. Fig. 7.2 shows Foucault's arrangement for measuring the speed of light. Light from an illuminated slit A, whose length is perpendicular to the plane of paper, is reflected from 45° glass plate P and rendered very nearly parallel by lens L. It is then reflected by a plane mirror R, and is focussed on a concave mirror C placed at a distance D, which is a *few metres*. R can rotate about an axis perpendicular to the plane of paper. This axis of R passes through the centre of curvature of the mirror C ; hence the rays reflected from mirror R towards C retrace their path on reflection at C.

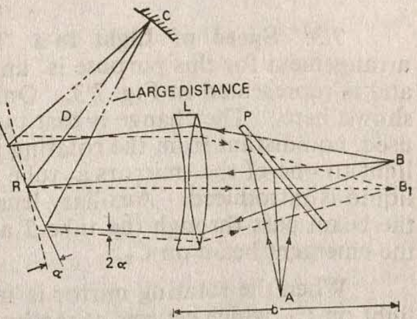


Fig. 7·2.
Optical setting of Foucault's method

If R is not rotating, the rays returning from C retrace their path, and a major part of it is transmitted to form an image B. If R rotates by angle α in the time taken by light in travelling the distance $2D$, the return beam is deflected by angle 2α with respect to the onward path. Hence the image is formed at B_1; the displacement $BB_1=s$ is measured by means of an eyepiece. It equals $f.\,2\alpha$ where f is the focal length of lens L.

If mirror R makes n rotations/sec, we have

$$s=f.2\alpha=2f.2\pi n\times2D/c$$

This gives

$$c=\frac{8\pi nf.}{s}D \qquad \ldots(7{\cdot}8)$$

With $D\sim20$ metres, the shift s obtained by Foucault was about 1 mm, the measurement of which could be accurate to about 1%. Increase of D would cause lowering of intensity ; for, if aperture of mirror C is d, then during the rotation of R the mirror C receives light for only a fraction $d/2\pi D$ of time in each rotation.

Example 7·2. Consider Foucault's experiment, using 14 transits over a distance D arranged by using mirrors causing forward and backward reflections. If speed of rotation of the mirror is 10 rev/sec., and focal length of the lens system is 50 cm, deduce the D needed to get a 1 mm displacement of the final image.

Solution.

$$\Delta t=\frac{14\times D(\text{in meters})}{3\times 10^8 m/\text{sec}}=\frac{14D}{3}\times 10^{-8}\ s$$

rotation of the mirror, $\alpha=2\pi n\,\Delta t$ radian$=\frac{28\pi D}{3}\times 10^{-7}$ radian

Shift of image 0·1 cm $=2\pi f=\frac{28\pi D}{3}\times 10^{-5}\Rightarrow D=3{\cdot}8$ m. ⊙

7·3. Speed of Light in a Transparent Liquid. Foucault's arrangement for this purpose is an adaptation of that in Fig. 7·2 and is represented in Fig. 7·3. Only the chief rays of beams are shown here. The change is that two concave mirrors C_1 and C_2 are used, equidistant from the rotating plane mirror R. In the path of light to one of the mirrors a tube T containing the experimental liquid is introduced. Auxiliary lenses L_1 and L_2 are needed to make the beam pass through the tube T as a parallel beam and to focus the emergent beam on C_2.

When the rotating mirror is in position R_1 it sends a flash of light on C_1, which returns at the instant when mirror is in position R_1'. The return beam gets focussed at B_1, where BB_1 is the displacement. When the mirror is in position R_2, it sends a flash of light on C_2, which returns at the instant when mirror is in position R_2'. The return beam therefore focuses at another point B_2, since time of flight is changed.

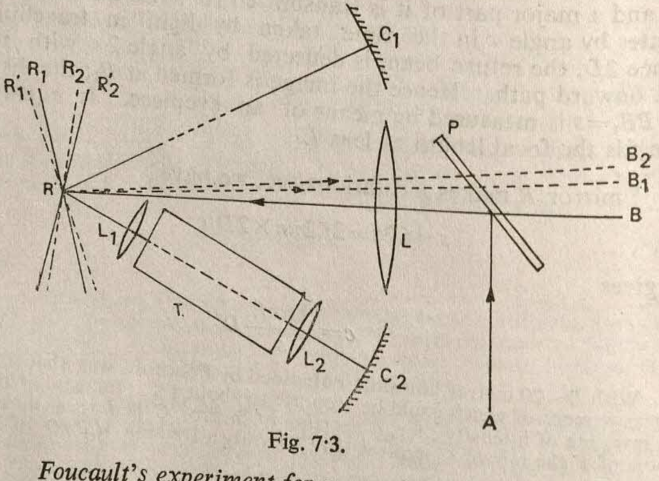


Fig. 7·3.

Foucault's experiment for measuring speed of light through a liquid

Let $RC_1=RC_2=D$. In path RC_2 the tube T, containing the liquid, occupies almost the entire length. Therefore, if c' is the speed of light in the liquid, we have

$$\frac{BB_2}{BB_1} \equiv \frac{s'}{s} = \frac{c}{c'}, \qquad \text{...(7·9)}$$

The measurement, when first made by Foucault (for water), was not very accurate. Yet it was of historic significance. He found $s' > s$, so that c' was less than c. Speed of light in water was less than its speed in air. The result gave a death blow to the corpuscular theory of light which predicted $c' > c$. Further, the ratio c/c' was found to be equal to refractive indexof the liquid, in *quantitative* support of the wave theory (see § 10·5).

7·4. **Michelson's Method.** Fig. 7·4 describes the arrangement used in the later experiments (1924-27) by Michelson in the famous Mount Wilson measurement of the speed of light. *The emphasis is on precision.* M_1 is an octagonal (eight-faced) mirror which can be rotated about its axis. Light from a slit source A falls on one face of M, reflects to a mirror M_2 and from there to another mirror M_3. Lens L is so set that it froms the focussed image of A on this mirror M_3. Mirror M_3 is a little above the plane of the diagram and M_2' a little below. Both lie in the focal plane of a concave mirror M_4 of large radius of curvature. Hence the beam from M_3 is rendered parallel by M_4, and goes to the *very distant* identical concave mirror M_5. The plane mirror M_6 at the focus of M_5 returns the beam on almost the same path†, so that the return beam is focussed by M_4 on M_3'.* From here light is reflected to M_2', thence to the opposite face of M_1 and is finally focussed by lens L' and the image B is seen with an eyepiece.

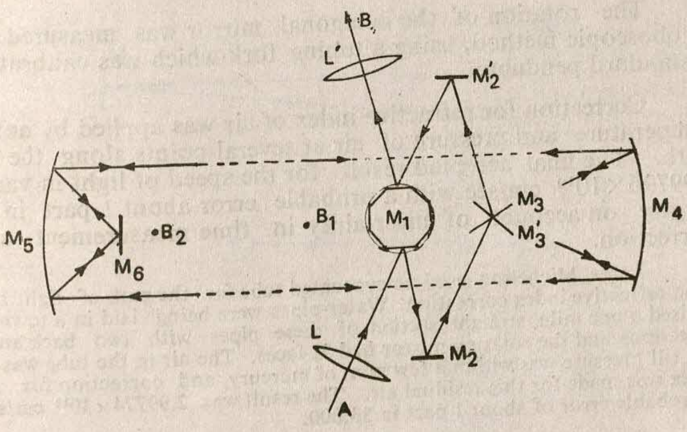


Fig. 7·4.

Ray path for Michelson's octagonal mirror method

---

† except for left right interchange.

* It needs only a slight tilt of M_4 to make the return beam focus not on M_3 but on M_3'.

The unit shown on the right hand side of Fig. 7·4 was on Mount Wilson, while the unit M_5, M_6 was on a 22 mile distant mountain.

The principle and theory is simple. If the octagonal mirror rotates at a constant speed such that in the time taken by light in travelling from one face of M_1 to the other (through the path just described) the octagonal mirror makes one-eighth rotation, the image B should be in the same position as when M_1 is not rotating at all. Otherwise, the image shifts to left or right. The speed of rotation of M_1 has to be adjusted to secure the null displacement state. If the distance M_1 to M_6 (through the path M_2, M_3, M_4, M_5) is D, and the octagonal mirror makes n rotations per second when the null shift of the image at B is achieved.

$$c=\frac{\text{distance}}{\text{time}}=\frac{2D}{1/8n}=16nD \qquad \ldots(7{\cdot}10)$$

The distance D was measured in two parts. Two marks B_1, B_2 were provided on two pillars at the two distant stations; their distance was measured by the Geodetic Survey of U.S.A. The accepted accuracy of this was 1 part in 200,000. Rest of these distances were measured by Michelson; since these are, after all, small additions to the large distance B_1 B_2, even a measurement correct to ±5 cm was enough*.

The rotation of the octagonal mirror was measured by the stroboscopic method, using a tuning fork which was calibrated with a standard pendulum.

Correction for refractive index of air was applied by noting the temperature and pressure of air at several points along the 22 mile path. The final accepted result for the speed of light in vacuum is 2.99796×10^{10} cm/sec with a probable error about 1 part in 10,000 largely on account of uncertainty in time measurement and air-correction.

Later, Michelson used an evacuated tube for the path of light to minimise refractive index correction. Water-pipes were being laid in a town and he utilised a one mile straight section of these pipes with two back and forth reflections and the rotating mirror had 32 faces. The air in the tube was pumped out till pressure was within a few mm. of mercury, and correction for refrative index was made for this residual air. The result was 2.99774×10^{10} cm/sec with a probable error of about 1 part in 30,000.

It may be remarked that this experiment was actually completed by Pease and Pearson after Michelson's death in 1931.

7.5. The Kerr Cell Method.** The method was first devised in 1925 by *Karolus* and *Mittelstaedt*, but we shall describe *Anderson's* form of it.

---

*Notice that 5 cm error in ∼30 km amounts to 1 part in 600,000

**This may be taken up after Polarization (Chapter 16).

In Fig. 7.5 L is a source of light, N_1 and N_2 make a pair of crossed nicols (Chapter 16), and KC is the Kerr cell, which is simply a pair of parallel plates immersed in nitrobenzene. This liquid is ordinarily isotropic, but it becomes anisotropic when an electrical field is applied to it. With no electric field across KC the nitrobenzene just passes the light polarized by N_1 as such and it is cut off by the *crossed* analyzer N_2. But when an electric field is present, the anisotropic nitrobenzene changes the nature of polarization of light passing through it, so that some light is able to pass the crossed analyzer. In short, the system has zero transmission at zero electric field and finite transmission when an electric field is present. A sinusoidal variation of electrical field gives sinusoidal variation of the transmitted light.

If the beam falls on a photo-cell a fluctuating current is produced. An amplifying system ensures that the constant background, if any, is cut off and only the fluctuating current is measured. The frequency of this equals the high frequency of the alternating voltage applied on the Kerr cell.

Now the mirrors M_1 to M_6 and the photo-cell R in Fig. 7.5 act as follows : Light from the lamp L is split by a half-silvered plate M_1. One part goes to mirror M_2 and is reflected back, part of it going to the photo-electric cell R. The other part goes to M_3 and then to the concave mirror M_4, which renders it parallel. The light

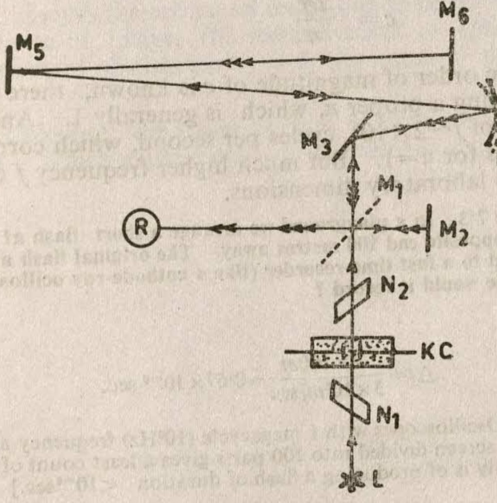


Fig. 7·5.
Anderson's method for measuring speed of light

then goes to plane mirrors M_5, M_6 and then retraces the path to M_1, part of it reaching R ultimately. An alter- native mirror M_4' can be brought in the position of M_4; its tilt and focal length are such that it returns the beams from M_3 on its own path, so that the path traversed is now smaller than that of the previous case by twice the length M_4, M_5, M_6.

When both the beams of light, with sinusoidal intensity variations produced by the Kerr cell, reach R the response depends on the time difference in the unequal paths of the two beams. Let T be the period of the sinusoidal variation. If the two beams reach R with time difference O, T, $2T$, etc., the photo-current fluctuates heavily, and hence the amplified current is large. If the two beams reach R with time difference $T/2$, $3T/2$, etc., the photo-current does not fluctuate appreciably, and hence the amplified current is low.*

In any experiment we first set M_4' in position and adjust M_2 (micrometer motion provided) to get minimum reading in the meter. Then we set M_4 in position and adjust M_2 again to get the minimum reading in the meter. The additional distance in the second case is M_4 to M_6 and back ± twice the motion of M_2 (other distances are common). Let this additional distance be D. Then time taken by light in traversing this distance is D/c and equals an integral multiple of the period of the oscillator of Kerr cell. If oscillator frequency is f, we get

$$\frac{D}{c}=nT=\frac{n}{f} \qquad (n, \text{ integer})$$

or

$$c=\frac{Df}{n} \qquad \ldots(7{\cdot}15)$$

Since the order of magnitude of c is known, there is no difficulty in assigning a proper n, which is generally 1. Anderson used an oscillator of $f \sim 2\times10^6$ cycles per second. which corresponds to $D \sim 150$ metres for $n=1$. But much higher frequency f can be used, reducing D to laboratory dimensions.

Example 7·3. In a playground we arrange a short flash at one end, and a mirror at the opposite end 100 metres away. The original flash and the return flash are both fed to a fast time-recorder (like a cathode-ray ocilloscope). What difference of time would it record ?

Solution.

$$\Delta t=\frac{2\times100m}{3\times10^8 m/\text{sec}}=0{\cdot}67\times10^{-6}\text{ sec.}$$

[Note. Oscilloscopes with 1 megacycle (10^6Hz) frequency are now easily available. Their screen divided into 100 parts gives a least count of $\cdot01\times10^{-6}$ sec. The problem really is of producing a flash of duration $<10^{-6}$sec.]

⊙

---

* When the photocell receives maximum intensity from one beam at the same instant as it receives maximum intensity from the other, the beams are said to be in agreement'. The statement has nothing to do with wave theory or any particular theory of light, but only with intensity variation. The period T just equals $1/f$, where f is the frequency of the oscillator connected to the Kerr cell.

PROBLEMS

7.1. In Fizeau's toothed wheel could the width of open and opaque spaces be kept unequal ? Could we use a rotation speed 10 times that which Fizeau used ?

7.2. In an experiment with Fizeau's method, using a 500-tooth wheel, the first 'eclipse' occurred at a speed of 1500 revolutions per minute. Calculate (*i*) the distance between the wheel and the distant mirror, (*ii*) the speed at which the next 'eclipse' would occur.

7·3. Comment on the relative merits of Fizeau's and Foucault's methods for measuring the speed of light.

7.4. A slit source of light, 0·1 mm in width, is placed at the focus of an *ideal* convex lens of $f'=50$ cm. Deduce the angle between the parallel beams corresponding to the two edges of the slit, and the amount of lateral spread they would show in travelling 20 km distance. Discuss the relevance of this for the intensity consideration in Fizeau's experiment.

7·5. In measuring the speed of light through a medium the 'comparison' method shown in Fig. 7·3 is preferred over direct measurement. Comment.

7·6. In Michelson's method of measuring the speed of light in vacuum, discuss the errors in result due to errors in (*i*) judging the coincidence of image, (*ii*) measurement of speed of rotation, (*iii*) measurement of distance *D*, and (*iv*) correction for refractive index of air.

7·7. Calculate the speed of light in a medium of refractive index 1·643. Which experiment is best suited to measure the speed and why ?

8

The Electromagnetic Theory of Light

When we say that *light* travels from place A to place B, the natural question is 'what' is this light made up of? In the case of sound the answer is that sound is made up of vibrations of the medium, which are transferred from particle to particle in the medium. These vibrations travel from A to B; the medium does no travel. This mechanism we call waves. What about light?

Various theories have been put forth about the nature of light. Among them the most important are the following:

(*a*) *Newton's corpuscular theory,*
(*b*) *Huygens' and Fresnel's ether wave theory,*
(*c*) *Maxwell's electromagnetic wave theory,*
(*d*) *Quantum electrodynamic theory.*

Experimental evidence is the sole basis for acceptance, rejection or modification of any theory. We shall discuss the evidence in later chapters. However, at this stage we shall make a brief survey of the theories.

8·1. Newton's Corpuscular Theory. This theory, put forth by Newton in 1675, considers light as *a stream of small particles (corpuscles) emitted from a luminous source and travelling at great speed.* Newton explained the various optical phenomena at that time as follows:

Colour. Newton assumed that corpuscles of different sizes created different colour effects on the retina of the eye.

Since there was no way of measuring the "size" of the corpuscles, no quantitative check could be applied on this assumption.

Rectilinear Propagation. At Newton's time no measurement of speed of light had been available, so one could assume infinite speed. Then rectilinear propagation follows naturally. From the now known value of c one can compute that in a horizontal travel of 1 mile the vertical fall under gravity would be less than 10^{-8} inch.*

---

* Note that smallness of mass is *not* involved in this calculation. It is the largeness of speed that limits the deflection to such a low value.

Departures from rectilinear propagation had not been observed for light at that time.

Refraction. To explain refraction Newton assumed an *ethereal medium*—far rarer, subtler and more elastic than air—pervading all space including pores of all bodies. The density of this 'ether' was assumed to vary from material to material. According to Newton refraction occurred because a denser medium (for ether) attracted the corpuscles, consequently bending their path towards the normal and increasing the speed of the corpuscles. Snell's law ($\sin i/\sin r =$ constant) could be proved on this basis, the relation being $n=$ speed in medium/speed in vacuum$=c'/c$.

Since measurements of even c were not available in Newton's time (see chapter 7), there was no way of checking whether n equals c'/c. Measurements now available show that n equals c/c', not c'/c.

Reflection. This was explained by Newton by assuming a force of repulsion at the *surface* of separation of the media. The force reversed the normal component of velocity of the corpuscle, while the parallel component remained constant, which fact led to the law of reflection.

We know that at the same boundary partial reflection and partial refraction occurs *simultaneously*. Hence one has to assume *attraction* to explain refraction and *repulsion* to explain reflection at the same boundary. Newton's answer to this problem is worthy of note. He stated that the *ether at the boundary was set into vibrations by the impinging corpuscles so that the ether entered into "fits of easy reflection and fits of easy refraction*."

It is worth recalling that Newton was the person who deduced the formulae for speed of longitudinal elastic waves in material media and speed of transverse waves on strings. And he also introduced the concept of the *ethereal medium* for light. But he used the ether concept to explain the phenomena at the boundaries only. Incidentally, Newton himself had discovered another phenomenon when the medium between two boundaries was very thin. That is 'Newton's Rings'. (§ 12·4). Here the "fits" varied as the thickness of the thin film varied. In an attempt to explain these alternate bright and dark rings by alternate "fits", Newton came very close to the wave theory.

The coming up of quantum theory is sometimes referred to as the revival or return of the corpuscular theory. This is not correct. Newton's corpuscles were material particles, subject to gravitation and also to attractive and repulsive forces at the boundaries of surfaces. His corpuscles of a given colour would change energy when their speed varied from medium to medium. The 'quanta' have entirely different laws governing them and wave-character is an integral part in so far as frequency is put in the equation $\varepsilon=h\nu$ for the energy of a 'quantum'.

8·2. Huygens' and Fresnel's Ether. That light may consist of a periodic disturbance transmitted through a medium in the form of waves, was first proposed by Huygens in 1678.* He assumed a medium, *luminiferous ether*, pervading all space (including vacuum) and having very high elasticity. A source of light created periodic disturbance in the ether, and this disturbance travelled as a wave, just as sound waves travelled in air.

---

* That is just 3 years after Newton propounded the corpuscular theory.

Huygens explained refraction of light satisfactorily but his prediction that light travelled slower in denser media was not verified experimentally for 170 years (Foucault, 1850). He could explain reflection too in a simpler way. As regards simultaneous occurrence of refraction and reflection, it was easier to understand on the wave theory, although their relative magnitudes could not be calculated for the next 150 years (Fresnel, 1822).

Diffraction and interference of light were not known at that time. Hence Huygens' wave theory got little experimental support. Further, Huygens could not explain rectilinear propagation on the wave theory. He also discovered polarization of light in 1690, but it became a hindrance, not help, to his wave theory, because he had *longitudinal* waves in mind, which could not explain polarization.

In retrospect, Huygens appears to have worked on the wave theory more on intuitive conviction than on experimental evidence, which came in sufficient measure only about 150 years later.

Fresnel, around the year 1820 *A.D.*, introduced the transverse ether wave theory. For the speed of mechanical waves through a medium the equations established by Newton are :

$$c\ (\text{longitudinal}) = \sqrt{\frac{k+(4/3)n}{\rho}} \qquad \ldots(8{\cdot}1a)$$

$$(c\ \text{transverse}) = \sqrt{\frac{n}{\rho}} \qquad \ldots(8{\cdot}.b)$$

where k and n are volume elasticity and rigidity respectively, and ρ is density of the medium. For lumineferous ether Huygens had taken $n=0$, but Fresnel took finite n in order to transmit transverse waves. He took $k \gg n$ in order to make c (longi)→∞, so that these waves did not participate in optical phenomena.

Fresnel could explain interference, diffraction, rectilinear propagation, polarization and double refraction on his theory. Quantitative evidence had been established (much of it by Fresnel himself) in support of this theory, a remarkable one being about the partial reflection and partial refraction of waves of different polarizations at the boundary of media.

Now, if all space is filled with ether (of finite rigidity) how do the planets move round without retardation ? Fresnel bypassed this question by assuming n to be vanishingly small (though finite), and saying that ρ was even more vanishingly small, so that the *ratio*/ρ was still as large as c^2.

Towards the end of the nineteenth century another question of great importance was raised. It ether stationary relative to the earth or relative to the sun or relative to a star ? Either assumption led to serious difficulties and conflicted with experiments. Hence the concept of ether, or any 'medium' for waves of light, had to be given up (see Chapter 24).

It is to be noted that although the *ether wave* concept is now dead, the valuable work of Fresnel and others on waves as such is applicable in the electromagnetic wave theory. The reason lies in an important principle : *the same equations have the same solutions*. By changing the meaning of symbols in an equation the interpretation changes but not the solution.

8·3. Maxwell's Electromagnetic Waves. A static electrical charge produces an electric field (Coulomb's law), a moving charge (*i.e.*, current) produces a magnetic field (Ampere's law), and an accelerated charge (*i.e.*, changing magnetic field) produces an electric field (Faraday's law). Also there are no free magnetic poles. Maxwell in the year 1873 considered these relationships and asked : if there is an oscillatory electrical circuit at station A, what happens at a distant station B ? His theoretical study showed that the oscillations *travel* to B with a speed whose *calculated* value is equal to the speed of light. In Maxwell's words, "*We have strong reason for believing that light is an electromagnetic phenomenon.*

Assuming that the reader has some background of electromagnetism and vectors, we outline Maxwell's theory below. Rationalised MKS units are used, with symbols E, D, H, B, i and ρ having their usual meanings.

(a) Equations of the Electromagnetic Field. The relations known at Maxwell's time were

(i) $\text{div}\ \mathbf{D}=\rho$, (iii) $\text{curl}\ \mathbf{H}=i$

(ii) $\text{div}\ \mathbf{B}=0$, (iv) $\text{curl}\ \mathbf{E}=-\partial\mathbf{B}/dt$

The first is Gauss's law, arising from Coulomb's law. The second expresses that isolated magnetic poles are absent. The third expresses Ampere's law with i as current density due to motion of free charge. The last one is Faraday's law.

Maxwell's critical analysis showed that Ampere's law was incomplete. In the space between the plates of a capacitor in an a.c. circuit there is no conventional current ($i=0$), yet curl $\mathbf{H}\neq 0$. Maxwell therefore included a term $\mathbf{dD}/dt$, apart from the conventional current i, in this equation (iii). This is called the *displacement current* and in free space and in a dielectric medium this is the only current present.

Thus, Maxwell's equations for the electromagnetic field in a dielectric (non-conducting) medium (where $\rho=0$) become

$$\text{div}\ \mathbf{D}=0 \tag{8·2}$$

$$\text{div}\ \mathbf{B}=0 \tag{8·3}$$

$$\text{curl}\ \mathbf{H}=+\partial\mathbf{D}/dt \tag{8.4}$$

$$\text{curl}\ \mathbf{E}=-\partial\mathbf{B}/dt \tag{8.5}$$

(b) Solution for an Isotropic Medium. The relations $\mathbf{D}=\varepsilon\mathbf{E}$ and $\mathbf{B}=\mu\mathbf{H}$ have the permittivity ε and permeability μ as tensors in general. But in an isotropic medium these are simple coefficients.

Eq. (8·5) now gives

$$\text{curl curl } \mathbf{E} = -\mu \text{ curl } \partial \mathbf{H}/dt$$

But curl $\partial \mathbf{H}/\partial t = (\partial/\partial t)$ curl **H**, and from Eq. (8·4) it equals $\varepsilon \partial^2 \mathbf{E}/\partial t^2$. That leads to

$$\text{curl curl } \mathbf{E} = -\mu\, \varepsilon\, \partial^2 \mathbf{E}/\partial t^2$$

The left side equals grad div $\mathbf{E} - \nabla^2 \mathbf{E}$ as an identity for all vectors. In our case div $\mathbf{E}=0$ (Eq. 8·2). So the result is

$$\nabla^2 \mathbf{E} = \mu\varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \qquad \ldots(8{\cdot}6)$$

This expresses the *space variations* of **E** in terms of the *time variations* of **E**.

Eq. (8·6) has the standard wave solution

$$\mathbf{E} = f\left[t - \frac{lx+my+nz}{v}\right] \qquad \ldots(8{\cdot}7)$$

where f is any function, and l, m, n are the direction cosines of normal to the wavefront travelling with speed v. Eq. (8·7) would give

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} = f''\left[t - \frac{lx+my+nz}{v}\right]$$

and

$$\nabla^2 \mathbf{E} \equiv \frac{\partial^2 \mathbf{E}}{\partial x^2} + \frac{\partial^2 \mathbf{E}}{\partial y^2} + \frac{\partial^3 \mathbf{E}}{\partial z^2}$$

$$= \frac{l^2+m^2+n^2}{v^2} f''\left[t - \frac{lx+my+nz}{v}\right] = \frac{1}{v^2}\, f''\left[t - \frac{lx+my+nz}{v}\right]$$

recalling that $l^2+m^2+n^2=1$. Substituting in Eq. (8·6), we find that Eq. (8·7) satisfies Eq. (8·6) if

$$v = 1/\sqrt{\mu\varepsilon} \qquad (8{\cdot}8)$$

In essence the theory shows that the time variations of E should travel in space with the speed determined by (Eq. 8·8), the parameters μ and ε being known from electromagnetism.

We could as well eliminate E (and D) between Eqs. (8·4) and (8·5) to deduce

$$\nabla^2 \mathbf{H} = \mu\varepsilon \frac{\partial^2 \mathbf{H}}{\partial t^2} \qquad \ldots(8{\cdot}9)$$

This would mean that the *time variations* of **H** travel in space with the same speed as the time variations of **E**.

For vacuum $\mu=\mu_0=1{\cdot}26\times10^{-6}$ henry/m and $\varepsilon=\varepsilon_0=8{\cdot}85\times10^{-12}$ farad/m. Eq. (8 8) then leads to $v=3{\cdot}0\times10^{8}$ ms^{-1}, which is equal to the speed of light in vacuum, c. For a dielectric medium μ is close enough to μ_0 and $\varepsilon = K\varepsilon_0$, where K is the dielectric constant. Hence speed of e.m. waves in a dielectric medium is given by

$$c' = 1/\sqrt{\mu_0 K \varepsilon_0} = c/\sqrt{K} \qquad \text{...(8.10)}$$

In practice c/c' gives the refractive index n and Eq. (8·10) leads to

$$n = \sqrt{K} \qquad (8\cdot11)$$

This relation is found to hold widely. Notice that K is not an arbitrarily introduced quantity, but one that is available from electrical experiments. Since K is frequency dependent, only the high frequency value of $\sqrt{K}$ would match the observed n.

(c) **Transverse nature of the wave.** Using $D = \varepsilon E$ and expanding *div E*, Eq. (8·2) for an isotropic medium will be

$$\varepsilon\left[\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}\right] = 0$$

Now consider a plane wave travelling along the x-direction. It means that E (also H) is independent of y and z. That leads to

$$\frac{\partial E_x}{\partial x} = 0 \text{ everywhere } \Rightarrow \frac{\partial^2 E_x}{\partial x^2} = 0$$

The x-component of Eq. (8·6) then gives

$$\frac{\partial^2 E_x}{\partial t^2} = 0 \quad \Rightarrow \quad E_x = \mathrm{a} + bt$$

This means that E_x is either zero ($a = b = 0$), or invariant with time ($b = 0$), or increasing uniformly with time ($b \neq 0$). The last two would not be part of a wave motion. Hence we conclude that $E_x = 0$ for a wave travelling along the x direction. Similar result can be established for H_x.

Thus a plane wave has no component of **E** or **H** along the direction of propagation. Eiectromagnetic waves are therefore transverse.

(d) **Relation between E and H.** Consider a plane wave travelling along the x-axis, with **E** vector along y-axis :

$$\mathbf{E} = E_y = f\left(t - \frac{x}{v}\right); \qquad E_z = 0$$

What is the magnitude and direction of **H** ? To deduce it we note that **H** is independent of y and z (wavefront parallel to yz plane) and also $H_x = 0$ (as proved earlier). That gives

$$\text{curl } \mathbf{H} = -\frac{\partial H_z}{\partial x} j + \frac{\partial H_y}{\partial x} k.$$

Now, using Eq. (8·4) leads to

$$-\frac{\partial H_z}{\partial x}j+\frac{\partial H_y}{\partial x}k=\varepsilon\frac{\partial E_y}{\partial t}j \qquad \text{...(8·12)}$$

Equating separately the coefficients of j and k gives us

$$\frac{\partial H_y}{\partial x}=0 \text{ and } -\frac{\partial H_z}{\partial x}=\varepsilon\frac{\partial E_y}{\partial t} \qquad \text{...(8·13)}$$

The first equation tells us $H_y=0$ (since $H_y=a$ constant over x is not a wave solution). We also have $H_x=0$. So **H** is along z-axis. It means that **H** and **E** are *mutually perpendicular*.

The second equation gives

$$-\frac{\partial H_z}{\partial x}=\varepsilon f'\left(t-\frac{x}{v}\right)$$

whence

$$H_z=v\,\varepsilon f=\frac{1}{\sqrt{\mu/\varepsilon}}\,E_y{}^{*} \qquad \text{...(8·14)}$$

and

$$B_z=\mu H_z=\sqrt{\mu\varepsilon}\;E_y=\frac{E_y}{v} \qquad \text{...(8·15)}$$

The ratio $\sqrt{\mu/\varepsilon}$ has the dimensions of impedence (ohms) and is called the *characteristic impedance* of the medium. For free space its value is ~ 380 ohm. Eq. (8·15) tells us that the variations of **E** and **B** are in same phase so that their magnitudes bear a constant ratio.

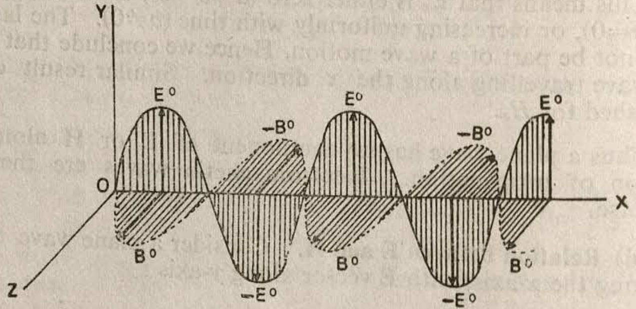


Fig. 8·1.

The **E** *and* **B** *vectors in an electromagnetic sinusoidal wave.* **E×B** *gives the direction of propagation* ; **E.B**=0 ; *and ratio* **E/B**=v *at all times, so that the maxima, minima and zeros of* **E** *and* **B** *occur simultaneously*, *that is* **E** *and* **B** *are in phase.*

---

* A simple way to remember this is that it leads to $\frac{1}{2}\varepsilon E^2=\frac{1}{2}\mu H^2$, the two sides representing the energy densities in the field due to E and H respectively.

(e) Invariance under Lorentz transformations. In Chapter 24 we will deal with Relativity : how events in one reference frame appear to an observer in uniform motion relative to the frame. If Galilean transformation of coordinates is used, Maxwell's relations (8·2)—(8·5) change their form. But if Lorentz transformation of space and time coordinates is used, these equations remain invariant. Conversely, if the transformation for keeping Maxwell's equations invariant is searched, then we reach Lorentz's form. We shall leave the matter till Chapter 24.

(f) Waves in anisotropic media and in conducting media. For anisotropic media ε and μ are to be treated as tensors, and we get the phenomenon of double refraction (Chapter 16). But we shall not go into the details thereof. Fresnel had earlier used the anisotropy of the elastic constants of luminiferous ether to get the same results.

For conducting media, Eq. (8·2) has ρ and Eq. (8·4) has conventional current i added to the right-hand side. That leads to a wave which is absorbed as it travels along. That explains why all metals (electrical conductors) are necessarily opaque. But we shall not give the details of the theory.

8·4. The Emission of Electromagnetic Waves. While Maxwell's treatment given above explains the nature of *propagation* of the e.m. waves in space, the nature of *emission* of the e.m. waves needs attention to the source. This follows better in a procedure adopted by Feynman. At a point distant r from a charge q we have the Coulomb field **E** (depending on r), and contributions depending on dr/dt and d^2r/dt^2. Feynman shows that the latter are taken into account by modifying the Coulomb's law to

$$\mathbf{E}\,(t) = -\frac{q}{4\pi\varepsilon_0}\left[\frac{1}{r'^2}\mathbf{e}_r{}' + \frac{r'}{c}\frac{d}{dt}\left(\frac{\mathbf{e}_r{}'}{r'^2}\right) + \frac{1}{c^2}\frac{d^2}{dt^2}(\mathbf{e}_r{}')\right] \qquad \ldots(8·16)$$

We explain the terms with Fig. 8·2. The charge q moving along $CQPD$ is at P at time t, and the observer is at O. The first term is the Coulomb field, but r is replaced by r', the position of Q at an earlier time $(t—r/c)$, called the *retarded time*; $\mathbf{e}_r{}'$ is the corresponding unit vector along QO. The second and third terms in Eq. (8·16) show contributions from velocity and acceleration of the charge, respectively, and each relates to the value at the retarded time.

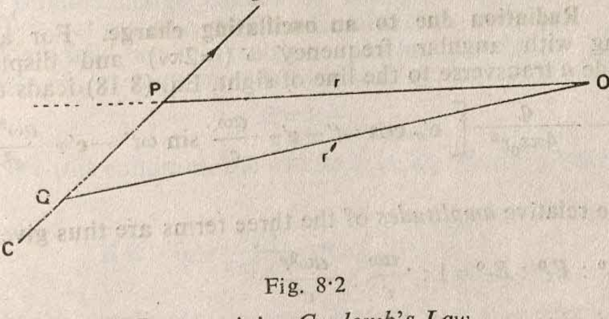


Fig. 8·2

Re-examining Coulomb's Law

We write Eq. (8·16) without proof; but we may state that c here equals $1/\sqrt{\mu\varepsilon}$, and that the equation is rigorous. Some approximations now follow. For particle speeds $\ll c$ the *magnitude* change r to r' is much smaller than r. Also, the time derivatives of unit vector $\mathbf{e}_r'$ have contributions only from the transverse velocity v_T' and transverse acceleration α_T' of the particle, and these are given by

$$\frac{d}{dt}(e'_r)=\frac{1}{r}v_T' \text{ and } \frac{d^2}{dt^2}(\mathbf{e}_r')=\frac{1}{r}\alpha_T' \qquad \ldots(8·17)$$

Proofs of these are given in textbooks on vectors. What is important is to see that the contributions of the second and third terms in Eq. (8·16) are parallel to v_T' and α_T' respectively.

These simplifications reduce Eq. (8·16) to

$$\mathbf{E}(t)=-\frac{q}{4\pi\varepsilon_0 r^2}\left[\mathbf{e}_r'+\frac{1}{c}\,\mathbf{v}_T'+\frac{r}{c^2}\alpha_T'\right] \qquad \ldots(8·18)$$

We repeat that this simplification will not be valid if v is comparable to c, or if the distance r is too small. We now use this equation to discuss the radiation in different cases.

(a) **Radiation due to an accelerated charge.** While the first two terms in Eq. (8·18) vary as $1/r^2$, the third one varies as $1/r$. Hence at large distance the dominant contribution to the radiation field [recalling $\alpha'(t)=\alpha(t-r/c)$] is

$$\mathbf{E}(t)=-\frac{q}{4\pi\varepsilon_0 c^2 r}\alpha_T\left(t-\frac{r}{c}\right) \qquad \ldots(8·19)$$

The direction of $\mathbf{E}$ is the same as that of α_T at the retarded time. Magnitudewise E is proportional to α. A charged particle in (say) a circular motion *must* radiate out energy, so also a charged particle which is retarded or deflected.

(b) **Radiation due to an oscillating charge.** For a charge oscillating with angular frequency ω $(=2\pi\nu)$ and displacement ampltitude a transverse to the line of sight, Eq. (8·18) leads to

$$\mathbf{E}(t)=-\frac{q}{4\pi\varepsilon_0 r^2}\left[\mathbf{e}'_r\cos\omega t'-\mathbf{e}'_T\frac{a\omega}{c}\sin\omega t'-\mathbf{e}'_T\frac{a\omega^2 r}{c^2}\cos\omega t'\right] \qquad \ldots(8.20)$$

The relative *amplitudes* of the three terms are thus given by

$$E_1^0 : E_2^0 : E_3^0=1:\frac{a\omega}{c}:\frac{a\omega^2 r}{c^2} \qquad \ldots(8·20a)$$

As ω and r increase, the second and third terms tend to become significant or dominant.

(c) **Radiation due to an oscillating dipole.** Two charges q and $-q$ separated by distance d are said to form an electric dipole of strength $p=qd$. If $d=d_0+a_0 \cos \omega t$, we have an oscillating dipole $p_0 \cos \omega t$ $(p_0=qa_0)$ superposed on a permanent dipole qd_0. Fig. 8·3 shows the Coulomb fields E_q and E_{-q} at 0 due to the two charges. Their resultant E is transverse to the line of sight, and $E/E_q=d \cos \theta/r$. In essence that adds a $(1/r)$ factor in the Coulomb term, and changes the direction of E to the transverse one. Using $p_0=qa_0 \cos \theta$, relation (8·20) now becomes

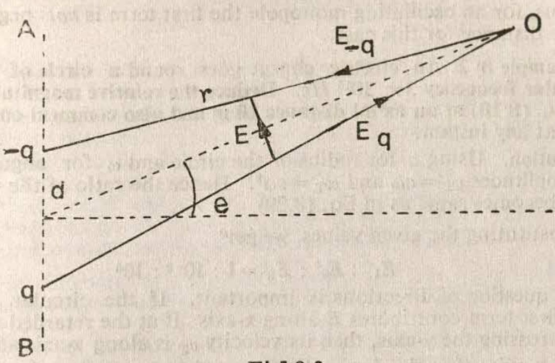


Fig 8·3

Deducing E for a Dipole.

$$\mathbf{E}\,(t)=-\frac{p_0}{4\pi\varepsilon_0 r^3}\mathbf{e}'_T\left[\cos \omega t'-\frac{\omega r}{c}\sin \omega t'-\frac{\omega^2 r^2}{c^2}\cos \omega t'\right] \quad ...(8·21)$$

The relative *amplitudes* of the three terms now become

$$E_1^\circ : E_2^\circ : E_3^\circ = 1 : \frac{\omega r}{c} : \frac{\omega^2 r^2}{c^2} \quad ...(8·22)$$

An isolated charge cannot be made to oscillate. Therefore Eq. (8·20) is not realistic, whereas Eq. (8·21) is. We now see that the Coulomb field becomes insignificant when $(\omega r/c) \gg 1$. Since $\omega=2\pi\nu$ and $c/\nu=\lambda$, we may put this condition as

$$2\pi r \gg \lambda \quad ...(8·23)$$

Under this condition the last term in Eq. (8·21) dominates, so that

$$\mathbf{E}\,(t)=\mathbf{e}'_T\,\frac{p_0\omega^2}{4\pi\varepsilon_0 c^2 r}\cos \omega\left(t-\frac{r}{c}\right) \quad ...(8·24)$$

Thus the amplitude is proportional to p_0 and ω^2 and falls off with r in the proportion $1/r$.

Exampl 8.1. An electric charge q oscillates simple harmonically with frequency $3\times10^{12} Hz$ and amplitude 10^{-10} m. Evaluate the relative magnitude of the three in Eq. (8·18) for an observer at distance 10^2 m in the equatorial plane.

Solution. Substituting the given values in Eq. (8·20) we get

$$E_1^\circ : E_2^\circ : E_3^\circ = 1 : 2\,\pi .\ 10^{-6} : 4\pi^2$$

Thus, for an oscillating monopole the first term is *not* negligible even at a 100metre distance for this case.

⊙

Example 8· 2. An electric charge goes round a circle of radius $10^{-10}\,m$ at an angular frequency $3\times\ 10^{15}\ Hz$. Deduce the relative magnitudes of the three terms in Eq. (8·18) at an axial distance 10 m and also comment on their relative directions at any instant.

Solution. Using a for radius of the circle and ω for angular frequcncy, we have amplitudes $v_T' = a\omega$ and $\alpha_T' = a\omega^2$. Hence the ratio of the three terms in Eq, (8·18) becomes same as in Eq. (8.20)

Substituting the given values, we get*

$$E_1^\circ : E_2^\circ : E_3^\circ = 1 : 10^{-3} : 10^5$$

The question of directions is important. If the circular path is in yz plane, the first term contributes E along x-axis If at the retarded time $t-r/c$ the particle is crossing the y-axis, then its velocity v_T is along z-axis and acceleration α_T is along y-axis : if at the time $t\text{-}r'/c$ the particle is crossing the z-axis then v_T is along y-axis and α_T is along z-axis. Thus the three terms contribute to E in three mutually perpendicular directions all the time.

⊙

Example 8·3 For an electric dipole oscillator of frequency $10^{12}\ Hz$ estimate the relative amplitudes of the position-dependent, velocity-dependent and acceleration-dependent components of the field at a distance 5cm in an equitorial plane.

Solution. Eq. (8·22) is useful here. Given $\nu = 10^{12}\ s^{-1}, r = 5$ cm $c = 3\times10^{10}$ cm s^{-1}, we have

$$\frac{\omega r}{c} = \frac{2\times3{\cdot}14\times10^{12}\times5}{3\times10^{10}} \approx 10^3$$

Hence $\qquad E_1^\circ : E_2^\circ : E_3^\circ = 1 : 10^3 : 10^6$

⊙

8·5. Energy Flux in an Electromagnetic Wave. If we consider a static electric field E, then the energy density in the space there is $\frac{1}{2}\varepsilon_0 E^2$. The simplest proof is through a parallel plate capacitor. With plates of area A and separation d, we have $C = \varepsilon_0 A/d$, $V = E.d$, and stored energy $\frac{1}{2}\ CV^2$. Hence energy density due to electric field is given by

$$W_E = \frac{\frac{1}{2}CV^2}{A.d} = \frac{1}{2}\varepsilon_0 E^2 \qquad \ldots(8{\cdot}25)$$

In an electromagnetic wave two things are different from a static field. The electric field is oscillating, $E_0 \cos \omega t$, so that mean E^2 has a value $\frac{1}{2}E_0^2$. Also, there is an accompanying magnetic field contributing to the energy density. Hence

$$\overline{W} = \overline{W_E} + \overline{W_H} = \tfrac{1}{4}\varepsilon_0 E_0^2 + \tfrac{1}{4}\mu_0 H_0^2 .$$

In view of Eq. (8·14) the two contributions are equal always. Hence

$$\overline{W} = \tfrac{1}{2}\varepsilon_0 E_0^2 \qquad \ldots(8{\cdot}25a)$$

Now, the electromagnetic wave travels with speed c. If we consider a plane wavefront of area s, then the energy in a cylinder of length $c\triangle t$ crosses this area in time $\triangle t$. The energy flow for unit area per unit time is called *energy flux* or *intensity* (I) of the radiation. Thus

$$I=\frac{W\times c\Delta t\times s}{\Delta t\times s}=Wc=\tfrac{1}{2}c\varepsilon_0 E_0^2 \qquad \ldots(8{\cdot}25b)$$

If E_0 is expressed in volt/m and c in m/s, we get I in joule $m^{-2}s^{-1}$.

If may be recalled that in elastic media the energy of waves is partly kinetic and partly potential. Their sum remains constant and the mean value of each is half of the total. The parallelism with the electromagnetic wave case is attractive. But it should not be carried too far. For instance, one must not try to deduce dE/dt and use its square to get the 'kinetic' part through $(dE/dt)^2$. In electromagnetic wave theory E itself contains $(2\pi\nu)^2$ term—not $2\pi\nu$—and thus the energy involves a frequency dependent term ν^4.

Another difference is that while kinetic and potential energy maxima at a given time are separated by $\lambda/4$, the maxima for W_E and W_H are in phase.

Example (8·4) Consider a 10 *kw* oscillator radiating isotropically. Calculate the amplitudes of *E* and *H* fields at 2 *km* distance from it.

Solution.

$$I=\frac{\text{Power}}{\text{Area}}=\frac{10^4\ Js^{-1}}{4\pi(2000\ \text{m})^2}=\frac{10^{-2}}{16\pi}Jm^{-2}s^{-1}$$

Equating this with $\frac{1}{2}c\varepsilon_0 E_0^2$ (with $1/4\pi\varepsilon_0=9\times10^9$ volt-m coul^{-1})

$$E_0^2=\frac{2I}{c\varepsilon_0}=\frac{10^{-2}}{c.8\pi\varepsilon_0}=\frac{10^{-2}\times9\times10^9}{2\times3\times10^8}(\text{volt}/m)^2$$

$$E_0\approx\mathbf{0{\cdot}40\ volt/m}$$

Using Eq. (8·14), with $\sqrt{\mu_0/\varepsilon_0}\approx380$ ohm we get

$$H_0=E_0/\sqrt{\mu_0/\varepsilon_0}\approx\mathbf{0{\cdot}10\times10^{-2}\ amp/m}.$$

⊙

8·6. Spatial distribution of dipole radiation. Fig. 8·4 shows an antenna AB connected to a high-frequency generator. The free

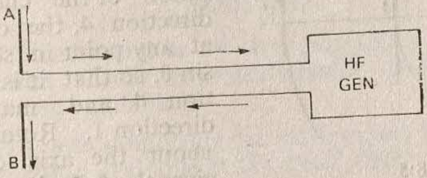


Fig. 8·4.

A high frequency generator connected to wires

electrons in wires A and B oscillate in unison. Hence in discussing the radiation field we may consider AB as a single wire in which a large number of electrons oscillate back and forth in unison. Their counterparts of opposite charge, *viz.*, the ion cores, remain fixed. For this reason the wire system A, B is called an *electric dipole oscillator*, although it really is a multitude of such oscillators acting *in unison*.

Example 8·5. A vertical antenna 2 m long carries alternating current of amplitude 200 ma at frequency 10 MHz. Deduce the amplitude of E in the radiation field at 100 m horizontal distance.

Solution. If a_0 is the displacement amplitude of each charge q, the velocity amplitude is $v_0=a_0 . 2\pi\nu$ and the electric current amplitude is

$$i_0=qv_0 N=(qa_0)\, 2\pi\nu . N=p_0 . 2\pi\nu N \qquad \ldots(8.26)$$

where N is the number of free electrons per unit length. The electric field amplitude in an equatorial plane is given by Eq. (8·24). Taking the fact that all electrons in 2 metre length oscillate is unison, we have

$$E^0=\frac{(2\pi\nu)^2 p_0}{4\pi\varepsilon_0 rc^2}\, 2N=\frac{2\pi\nu}{4\pi\varepsilon_0 rc^2}\, 2i_0 \qquad \ldots(8{\cdot}27)$$

using Eq. (8·26) to eliminate p_0N. Now we have

$$\nu=10^7 \text{ sec}^{-1},\ r=100 \text{ m},\ c^2=9\times10^{16}\ m^2\text{sec}^{-2},$$

$$i_0=2\times10^{-1} \text{ coul sec}^{-1}, \text{ and } (1/4\pi\varepsilon_0)=9\times10^9 \text{ volt. m. coul}^{-1}$$

Substitution in Eq. (8·27) gives

$$E^\circ=\frac{2\pi .\ 10^7 . 9\times10^9 . 4\times10^{-1}}{100\times9\times10^{16}}$$

$$=2{\cdot}5\times10^{-2} \text{ volt/m}$$

⊙

In Fig. 8·5, we imagine the dipole oscillator to be situated at the position O, with its axis along the direction 4. We draw spheres with O as centre, with radii sufficiently *large* so that Eq. (8.24) may be treated as valid. Since acceleration of the charges is along direction 4, the oscillating field at any point in space varies as $\sin\theta$, so that it is zero in direction 4 and maximum along direction 1. Rotating the figure about the axis of the dipole gives the full distribution of the field. The intensity varies as the square of the amplitude, so that we may write

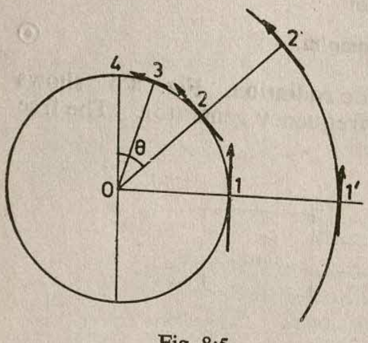


Fig. 8·5

Distribution of electric field due to the oscillating dipole

$$I\,(\theta, r)\propto\frac{\sin^2\theta}{r^2} \qquad \ldots(8{\cdot}25)$$

where θ and r define the point under consideration.

The electric dipole oscillator thus gives us all the essentials of the electromagnetic wave generation and propagation, including polarization and the directional dependence. From the metal wire dipoles driven by high frequency oscillators to the emission of light by atoms and molecules, and of γ rays by nuclei of atoms is a long step. But Maxwell's electromagnetic theory of light knits them together into one basic phenomenon : the *electromagnetic waves*.

8·7. Background of the Quantum Theory. One great advance in adopting the electromagnetic theory of light is that absorption, emission, scattering and all other interactions of light with matter can be explained along with (*not apart from*) the constitution of matter in terms of electrically charged particles and related chemicaland physical properties. Around the year 1900 A.D. it appeared that the electromagnetic wave picture could explain all the phenomena.

However, in quantitative details there came up some serious problems, which we will take up in Chapter 19. We find that light shows not only all the properties of waves, but somehow the energy and momentum are associated in bundles of size $h\nu$ and $h\nu/c$ respectively, where h is a universal constant, called Planck's constant. This was the quantum concept.

We shall resist the temptation of giving right now the thrilling story of discovery after discovery supporting the quantum concept. What is important for us is that a structure of mathematical analysis and philosophical concepts had to be built up to give us a comprehensive theory incorporating the wave and quantum aspects together. This is the field of *Quantum Electrodynamics* or the wider branch of Physics called *Quantum Mechanics* which has caused a revolution in our expectations of the future. We shall deal with it in Chapter 23.

In the meantime the reader will do well to understand the wave theory firmly, since that is the foundation on which to build up.

PROBLEMS

8·1. (*a*) What would be the strongest arguments against Newton's corpuscular theory if measurements of speed of light in vacuum alone were available ?

(*b*) Write a note supporting the wave theory (against Newton's) with the information available at Huygens' time.

8·2. In an experiment done to simulate Newton's explanation of refraction, two smooth horizontal planes, separated by height h, are connected with a smooth sloping board, leaving no sharp edges. A steel ball is rolled on the upper surface with speed v and varying angle θ with the normal to the boundary edge. The ball rolls down to the second surface and shows speed v' and angle θ'. Experiment then proves $v'/v = \text{constant} = \sin\theta/\sin\theta'$. Analyse the mechanics of the experiment.

8·3. A beam of light projected horizontally shows a 'fall' of less than 0·1 mm in travelling 10 km distance. What information do we deduce about the speed of light, assuming the corpuscular theory and gravitational effect on corpuscles as on material particles?

8·4. What assumptions about ether were used by Fresnel to account for transverse waves and absence of longitudinal waves ?

8·5. Name the laws expressed by Maxwell's equations div $D=\varrho$ and curl $\mathbf{H}=\mathbf{i}+\partial\mathbf{D}/\partial t$. State the reason for the $\partial\mathbf{D}/\partial t$ term in the second equation. Also write these laws for an isotropic dielectric medium in terms of $\mathbf{E}$.

8·6. (*a*) Show that for an e.m. wave travelling along z-axis, we have $\mathbf{E}_z=\mathbf{H}_z=0$.

(*b*) Show that for a plane e.m. wave $\mathbf{H}.\mathbf{E}=0$ and $\mathbf{H}=\mathbf{E}\sqrt{\mu/\varepsilon}$. State the *MKS* units of $\mathbf{H}$, $\mathbf{E}$ and μ/ε.

8·7. An electric charge has velocity $3\times10^6\ ms^{-1}$ and acceleration $5\times10^{12}\ ms^{-2}$, both at angle 90° from the line of sight. Deduce the relative magnitudes of the three terms in Eq. (8·18) for an observer at distance (i) 10 *m*, (ii) $10^6 m$.

8·8. Consider a star at distance 10^2 light-years from the earth, moving with velocity 0·80*c* normal to the line of sight. Comment on the 'retarded time' in dealing with the light from the star as seen from the earth, and deduce de_r'/dt.

8·9. An electron in Bohr orbit of *H* atom has radius 0·5A° and orbiting frequency $7\times10^{15} Hz$. Deduce the electric field amplitude at a distance 2 *cm* along the axis.

8·10. For an oscillating dipole the dominant field at a large distance is due to the acceleration dependent term. Establish the significance of '*large distance*', comparing the cases of frequencies 1*MHz* to 10^8 *MHz*.

8·11. If a plane polarized beam of light in air has intensity $40\ Jm^{-2}s^{-1}$, compute the amplitudes of *E*, *D*, *H* and *B*.

8·12. An antenna 0·80 *m* long has an oscillating current 2*ma* at frequency 20 *MHz*. Deduce the *E* and *H* field amplitudes at a distance 400 *m* at angle 30° from the length of the antenna.

9

Vibrations and Waves

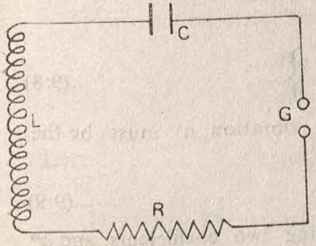
If you drop a pebble in a pond of water a *disturbance* is created at the point of fall. With passage of time this creates disturbances at successive positions of water. The process is called *wave motion.* The disturbed point of fall is the source and the water surface is the medium for the wave motion.

If a repeated disturbance is created at the source, there is a sustained wave in the medium. The simplest repeated disturbance is the *harmonic vibration*, and the wave created by it is the *harmonic wave.* But all practical oscillators have a finite damping, which we will examine. Also, all practical media show wave speed depending on frequency, which has important consequences. These we will consider.

The treatment of vibrations and waves becomes mathematically more convenient if we use complex quantities. In this chapter we will introduce this, side-by-side with the normal treatment.

Since the electromagnetic theory of light considers the source as an electrical oscillator, we will use that kind of oscillation, rather than a mechanical one. But much of the material in this chapter is applicable to all kinds of oscillators and waves.

9·1. Harmonic Oscillator. Consider the electrical circuit of Fig. 9·1 with $R=0$. Let us see what happens if the gap G is connected to a battery and thereafter closed. If at any time t the capacitor has charge q the voltage across it is q/C, and if the charge is changing at the rate dq/dt (the current), then the current is changing at the rate d^2q/dt^2. Faraday's law gives the *e.m.f.* across L as $-Ld^2q/dt^2$. Equating this *e.m.f.* with the voltage across the condenser gives



Fig, 9·1

A possible oscillating electrical circuit.

$$\frac{q}{C}=-L\frac{d^2q}{dt^2} \qquad \text{...(9·1)}$$

For convenience we introduce

$$\omega_o^2=\frac{1}{LC} \qquad (\ldots 9\cdot2)$$

to set Eq. (9·1) as

$$\frac{d^2q}{dt^2}+\omega_0^2q=0 \qquad \ldots(9\cdot3)$$

This equation has the general solution

$$q=A_c\cos\omega_0t+A_s\sin\omega_0t \qquad \ldots(9\cdot4)$$

as can be seen by substitution. The constants A_C and A_S are arbitrary, unless initial conditions are given. One can write Eq. (9·4) as

$$q=A\cos(\omega_0t+\phi) \qquad \ldots(9\cdot5)$$

where A and ϕ become the arbitrary constants.

Eq. (9·4) represents a harmonic oscillation of frequency $\omega_0/2\pi$. Amplitudes A_C and A_S for the cosine and sine contributions are represented separately here, whereas in Eq. (9·5) A is the total amplitude $[=(A_C^2+A_S^2)^{1/2}]$ and ϕ is the initial phase. Eq. (9.3) is the differential equation for the harmonic oscillator.

9.1 A. Use of complex amplitude

As an alternative, try for Eq. (9·3) the solution.

$$q=ae^{pt} \qquad \ldots(9\cdot6)$$

Substitution in Eq. (9·3) leads to $p=+i\omega_0$ or $-i\omega_0$. Hence the most general solution is

$$q=ae^{i\omega_0t}+a'e^{-i\omega_0t} \qquad \ldots(9\cdot7)$$

Here a and a' are the arbitrary constants. But unlike A_C and A_S (or A and ϕ) these a and a' are not real, but are complex quantities. Their relations with A_C and A_S are easy to see. Eq. (9·7) gives

$$q=(a+a')\cos\omega_0t+i(a-a')\sin\omega_0t$$

Comparison with Eq. (9·4) gives

$$\left.\begin{array}{r} a+a'=A_C \\ i(a-a')=A_S \end{array}\right| \Rightarrow \left.\begin{array}{l} a=\tfrac{1}{2}(A_C-iA_S) \\ a'=\tfrac{1}{2}(A_C+iA_S)=a^* \end{array}\right\} \qquad \ldots(9\cdot8)$$

The essence is that for Eq. (9·7) to represent a real vibration a' must be the complex conjugate of a. So we rewrite Eq. (9·7) as

$$q=ae^{i\omega_0t}+a^*e^{-i\omega_0t} \qquad \ldots(9\cdot9)$$

We repeat that a here is a complex quantity, involving two coefficients, and a^* is not an independent quantity. An interesting way of looking at Eq. (9·9) is that change $\omega_0\rightarrow-\omega_0$ means $A_S\rightarrow-A_S$, which means $a\rightarrow a^*$. Thus the two terms of Eq. (9·9) represent contributions of frequencies ω_0 and $-\omega_0$, and sum of the two terms is a real oscillation with $A_C=a+a^*$ and $A_S=i(a-a^*)$.

Solutions (9·4) and (9·9) are equally good. But the former deals with real amplitudes, separated for the cosine and sine contributions, while the latter deals with complex amplitudes, with cosine and sine contributions mixed in the complex form, needing the addition of a complex conjugate term of frequency $-\omega_0$. Now, we could as well leave the second term in Eq. (9·9) unwritten and state Eq. (9·9) as

$$q=ae^{i\omega_0 t} \qquad \text{...(9·10)}$$

The other term is left as understood. We often need to evaluate $\overline{q^2}$, and in the representation of Eq. (9·10) we define

$$\overline{q^2} \equiv |q^2| = qq^* + q^*q \qquad \text{...(9·11)}$$

This leads to $\overline{q^2} = 2aa^* = \frac{1}{2}(A_C^2 + A_S^2)$, the same as we get from Eq. (9·4)

The great advantage of the form (9·9) of describing the oscillation lies when we reduce it to Eq. (9·10) with the understanding expressed in Eq. (9·11).

Example 9·1. Find the sum of two oscillations (i) $A_1 \cos(\omega_0 t+\phi_1)$ and $A_2 \cos(\omega_0 t+\phi_2)$, (ii) $a_1 e^{i\omega_0 t}$ and $a_2 e^{i\omega_0 t}$.

Solution. (i) The sum amounts to

$$\begin{aligned} q &= [A_1 \cos\phi_1 + A_2 \cos\phi_2]\cos\omega_0 t - [A_1 \sin\phi_1 + A_2 \sin\phi_2]\sin\omega_0 t \\ &= A\cos(\omega_0 t + \phi) \end{aligned} \qquad \text{...(9.12)}$$

where

$$\left.\begin{aligned} A^2 &= A_1^2 + A_2^2 - A_1A_2\cos(\phi_1-\phi_2) \\ \tan\phi &= \frac{A_1\sin\phi_1 + A_2\sin\phi_2}{A_1\cos\phi_1 + A_2\cos\phi_2} \end{aligned}\right\} \qquad \text{...(9·13)}$$

You will see that several steps have been left out in the derivation.

(ii) In this case the sum is simply

$$q = a_1 e^{i\omega_0 t} + a_2 e^{i\omega_0 t} = (a_1+a_2)e^{i\omega_0 t}$$ ⊙

9·2. Damping in an Oscillator. If resistance $R \neq 0$ in the circuit of Fig. 9·1, there is a voltage drop $R\,dq/dt$ across it, so Eq. (9·1) becomes

$$\frac{q}{C} + R\frac{dq}{dt} = -L\frac{d^2q}{dt^2} \qquad \text{...(9·14)}$$

For convenience we write this as

$$\frac{d^2q}{dt^2} + \frac{1}{\tau}\frac{dq}{dt} + \omega_0^2 q = 0 \qquad \text{...(9·15)}$$

where

$$\omega_0^2 = \frac{1}{LC} \text{ and } \tau = \frac{L}{R} \qquad \text{.. (9·16)}$$

If we try the solution

$$q = ae^{pt} \qquad \text{...(9·17)}$$

Eq. (9·15) requires that

$$p^2 + \frac{1}{\tau}p + \omega_0^2 = 0$$

For $(1/\tau)^2<\omega_0^2$, which means $R<\sqrt{L/C}$, this leads to

$$p=-\frac{1}{2\tau}\pm i\sqrt{\omega_0^2-\frac{1}{\tau^2}}=-\frac{1}{2\tau}\pm i\omega. \qquad \ldots(9{\cdot}18)$$

where

$$\omega^2=\omega_0^2-\frac{1}{\tau^2} \qquad \ldots(9{\cdot}19)$$

Since p has two values, the most general solution of Eq. (9·15) is

$$q=e^{-t/2\tau}\,[ae^{i\omega t}+a'e^{-i\omega t}] \qquad \ldots(9{\cdot}20)$$

$$=e^{-t/2\tau}\,[(a+a')\cos\omega t+i\,(a-a')\,\sin\omega t]$$

Since q is a real function of time, both $a+a'$ and $i(a-a')$ must be real.* Calling them A_c and A_s respectively,

$$q=e^{-t/2\tau}\,(A_c\cos\omega t+A_s\sin\omega t) \qquad \ldots(9{\cdot}21)$$

$$=e^{-t/2\tau}\,A\cos(\omega t+\phi) \qquad \ldots(9{\cdot}22)$$

These may be compared with Eqs (9·4) and (9·5). Firstly, ω_0 changes to ω (Eq. 9·19), and secondly, the amplitude falls with time exponentially.

The energy ε of the oscillating circuit† is proportional to the (amplitude)^2, whence

$$\varepsilon=\varepsilon_e e^{-t/\tau} \qquad \ldots(9{\cdot}22)$$

The quantity $\tau=L/R$ is called the *relaxation time*. It is the time in which the *energy falls to* $1/e$ *of the original value.*

Another related quantity is the *quality factor* Q, defined as $2\pi\div$fractional energy loss per cycle. From Eq. (9·22) one readily gets

$$Q=\frac{2\pi}{T\left(\dfrac{1}{\varepsilon_0}\dfrac{d\varepsilon}{dt}\right)}=\omega\tau\approx\omega_0\tau \qquad \ldots(9{\cdot}23)$$

There is another way of defining Q. In resonance between a driver agency and the oscillator, the latter picks up maximum energy if the driver source has frequency equal to ω. If ω_1 and ω_2 are the limits at which the picked up energy is half of that at ω, then $\omega_2-\omega_1=\triangle\omega_{1/2}$ is called *resonance half-width*, and we define

---

*This means a' is complex conjugate of a (see Eq. 9·8),

†$\varepsilon=\dfrac{q^2}{2C}+\dfrac{1}{2}L\left(\dfrac{dq}{dt}\right)^2\approx\dfrac{q_0^2}{C}$, as you may see from $q=q_0\cos(\omega_0 t+\phi)$, with $\omega_0=1/\sqrt{LC}$. The damping term may be added separately.

$$Q\equiv\frac{\omega}{(\triangle\omega_{1/2})}\equiv\frac{\nu}{(\triangle\nu)_{1/2}}\equiv\frac{\lambda}{(\triangle\lambda)_{1/2}} \qquad \ldots(9{\cdot}24)$$

Theory shows that this relation also leads to $Q=\omega\tau$. Thus an oscillator with large τ (small damping) also gives sharp resonance peak.

In the case of visible light you have typical $\omega\approx10^{+15}\ s^{-1}$ and $\approx10^{-9}s$, leading to $Q\approx10^5$ from Eq. (9.23). In terms of Eq. (9.24) this will mean that for $\lambda\approx5000\ A^\circ$ you have $\triangle\lambda_{1/2}\approx0{\cdot}05\ A^\circ$.

You may shift attention from the oscillator to the field at a distant point. We have seen in Chapter 8 that the electrical field E is proportional to q, and energy flux (intensity) is proportional to E_0^2. Thus the intensity in the radiation field falls the same way as ε. Observations in the radiation field thus lead to Q of the source. Measurement of $\lambda/\triangle\lambda_{1/2}$ pertains to the field, from which ω and τ may be deduced.

Example 9·2. In an L, C, R series circuit, the values are $L=10\mu h$, $C=0{\cdot}20\mu f$, $R=1{\cdot}0\times10^{-2}$ ohm. Deduce the oscillation frequency, relaxation time, and Q value of the circuit.

Solution.

$$\omega_0^2=\frac{1}{LC}=\frac{1}{10\times10^{-6}\times0{\cdot}20\times10^{-6}}=0{\cdot}5\times10^{12}s^{-2}$$

$$\tau=\frac{L}{R}=\frac{10\times10^{-6}}{1\times10^{-2}}10^{-3}s$$

Since $\omega_0^2\gg1/\tau^2$, we have $\omega\approx\omega_0$. Hence

$$\text{frequency, } \nu=\frac{\omega_0}{2\pi}=1{\cdot}13\times10^5s^{-1}$$

$$\text{quality factor, } Q=\omega_0\tau=0{\cdot}71\times10^3.$$

⊙

9.2A. Use of a Complex Frequency

In Eq. (9·20), a' can be seen as a complex conjugate of a, so that the entire second term is a complex conjugate of the first term there. As before (see Eq. 9·10) we then write Eq. (9·20) as

$$q=ae^{i(\omega+ik)t} \qquad \ldots(9{\cdot}25)$$

leaving the second term as unstated [Here k is written for $1/2\tau$ for convenience.] The understanding is that the written term contains all the information and that q^2 is given by qq^*+q^*q, as in Eq. (9·11).

In Eq. (9·25) not only the amplitude a is complex, but the angular frequency $\omega+ik$ is also complex. The imaginary part ik represents absorption, and the real part ω represents phase change with time, *i.e.*, frequency. But Eq. (9·25) is a compact and useful way of expressing a damped oscillator.

Example 9·3. The electrical field in a certain region of space oscillates as $E=E_0\exp i(\omega+ik)t$, where $E_0=(5+4i)\ 10^{-3}$ volt/m. Deduce the real amplitude of E at $t=0$, initial phase, aud the energy in the field at $t=0$.

Solution. The given representation amounts to

$$E=E_0\exp[i(\omega+ik)t]+E_0^*\exp[-i(\omega-ik)t]$$

Simple substitution leads to (check it)

$$E=10^{-3}e\text{-}kt\,[10\cos\omega t - 8\sin\omega t]\ \text{volt/m}$$
$$=(12{\cdot}8\times10^{-3})e\text{-}kt\cos(\omega t+\phi),\ \text{where } \phi=\tan^{-1}(4/5).$$

At $t=0$, amplitude is $12{\cdot}8\times10^{-3}$ V/m. initial phase $=\tan^{-1}(0{\cdot}8)$

Energy density of the field is obtained from Eq. (8·28) as

$$\varepsilon=\tfrac{1}{4}\varepsilon_0\,|\,EE^*\,|=\tfrac{1}{4}\varepsilon_0\exp(-2kt)\,E_0E_0^*$$
$$=\varepsilon_0\exp(-2kt)$$

where
$$\varepsilon_0=\tfrac{1}{4}\varepsilon_0E_0E_0^*=\frac{41}{4}\varepsilon_0\times10^{-6}\,(V^{-1}m^{-1}\,\text{coul})(Vm^{-1})^2$$
$$=9{\cdot}1\times10^{-17}\ \text{Jm}^{-3}.$$

⊙

9·3. Fourier Series and Fourier Transform. Fourier's theorem is a theorem in pure mathematics, but is of great use in Physics. Consider a *periodic function* of variable t :

$$y=f(t)$$

with period T, *i.e.*, frequency $\nu=1/T$, or angular frequency $\omega_0=2\pi/T$. Here t (and T) need not be time, but any parameter ; as one instance it could stand for distance, in which case y is periodic (repetitive) in space, and T would then mean a spatial distance (like wave length).

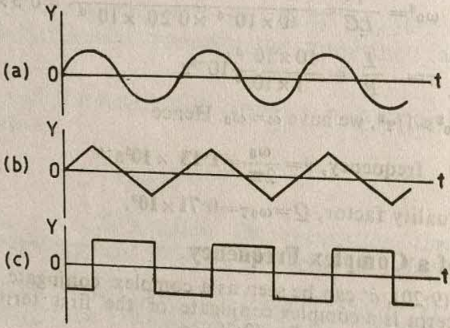


Fig. 9·2.

Three functions $y=f(t)$ of different forms but same period. How are they related ?

In Fig. 9·2 we have three periodic functions of the same period. How are they related ? Fourier's theorem gives the answer.

Function (a) in Fig 9·2 is harmonic and may be written as

$$y=A\cos(\omega_0 t)+B\sin(\omega_0 t) \qquad ...(9{\cdot}26)$$

Fourier's theorem states that (b) or (c) or any other *periodic function* of angular frequency ω_0 can be written as the series sum :

$$y=f(t)=A_0+\sum^{\infty}A_n\cos n\omega_0 t+\sum^{\infty}B_n\sin n\omega_0 t \qquad ...(9{\cdot}27)$$

The series in Eq. (9·27) is called Fourier Series. Further, the coefficients A_0, A_n, B_n may be deduced from the given $f(t)$ from the equations

$$A_0 = \frac{1}{T}\int_0^T f(t)\,dt \qquad \ldots(9·28\ a)$$

$$A_n = \frac{2}{T}\int_0^T f(t)\cos(n\omega_0 t)\,dt \qquad \ldots(9·28\ b)$$

$$B_n = \frac{2}{T}\int_0^T f(t)\sin(n\omega_0 t)\,dt \qquad \ldots(9·28\ c)$$

We shall not give the proof of Eq. (9·27), but the truth of Eq. (9·28) may be seen by simple substitution and integration.

9.3. A. Use of Exponential Representation

We recall that Eq. (9·4) could be put in the form of Eq. (9·9). For the n^{th} harmonic in the Fourier series (9·27) we can thus write

$$a_n e^{in\omega_0 t} + a_n^* e^{-in\omega_0 t}$$

where $a_n = \frac{1}{2}(A_n - iB_n)$ and a_n^* is the complex conjugate of a_n (See Eq. 9·8). Fourier series can then be written as

$$y = f(t) = \sum_{n=-\infty}^{\infty} a_n e^{in\omega_0 t} \qquad \ldots(9·29)$$

where the complex conjugate counterpart is included by using negative n values too. If you desire to get the amplitude a_m of the m^{th} harmonic, you have

$$\int_0^T f(t) e^{-im\omega_0 t}\,dt = \sum_{n=-\infty}^{+\infty}\int_0^T a_n e^{in\omega_0 t} e^{-im\omega_0 t}\,dt = T.a_m$$

because all integrals for $m \neq n$ vanish, and for $n = m$ the integral is just $\int_0^T a_m dt = a_m T$. Thus

$$a_m = \frac{1}{T}\int_0^T f(t)\, e^{-im\omega_0 t}\,dt \qquad \ldots(9·30)$$

Thus the Fourier series (9·27) and equations for the Fourier coefficients (9·28 *a, b, c*) simplify to Eqs. (9·29) and (9·30) respectively. We repeat that the coefficients a_m here are complex quantities.

Example 9·4. For the periodic oscillation given by $f(t) = +a$ for $0 < t < T/2$ and $f(t) = -a$ for $T/2 < t < T$, deduce the Fourier series.

Solution. Using Eqs. (9·28) we have

$$A_o = \frac{1}{T}\left[\int_o^{T/2} a\ dt + \int_{T/2}^{T} -a\ dt\right] = 0$$

$$A_n=\frac{2}{T}\left[\int_0^{T/2} a\cos\frac{2\pi nt}{T}dt+\int_{T/2}^{T}-a\cos\frac{2\pi nt}{T}\,dt\right]$$

$$=\frac{2a}{T}\cdot\frac{T}{2\pi n}\left[\left\{\sin\frac{2\pi nt}{T}\right\}_0^{T/2}-\left\{\sin\frac{2\pi nt}{T}\right\}_{T/2}^{T}\right]=0$$

$$B_n=\frac{2a}{T}\cdot\frac{T}{2\pi_n}\left[\left\{-\cos\frac{2\pi nt}{T}\right\}_0^{T/2}+\left\{\cos\frac{2\pi nt}{T}\right\}_{T/2}^{T}\right]$$

$$=\frac{a}{\pi n}\left[(-\cos\pi n+1)+(\cos 2\pi n-\cos\pi n)\right]$$

$$=\frac{4a}{\pi n}\text{ for odd } n\text{, and 0 for even } n.$$

The Fourier series is therefore

$$f(t)=\frac{4a}{\pi}\left[\sin\frac{2\pi t}{T}+\frac{1}{3}\sin\frac{6\pi t}{T}+\frac{1}{5}\sin\frac{10\pi t}{T}+\ldots\ldots\right]$$

Alternatively, using the complex representation, Eq. (9.30) gives—

$$a_n=\frac{1}{T}\left[\int_0^{T/2} a\,e^{-i2\pi nt/T}dt+\int_{T/2}^{T}-a\,e^{-i2\pi t/T}dt\right]$$

$$=\frac{a}{T}\cdot\frac{2}{-i2\pi n}\left[\left(e^{-i\pi n}-1\right)-\left(e^{-i2\pi n}-e^{-i\pi n}\right)\right]$$

$$=-i\frac{4a}{\pi n}\text{ for odd } n\text{, and 0 for even } n.$$

$$\therefore\qquad f(t)=\sum_{\text{odd } n}\left[-i\frac{4a}{\pi n}e^{i2\pi nt/T}+i\frac{4a}{\pi n}e^{-i2\pi nt/T}\right]$$

$$=\sum_{\text{odd } n}\frac{4a}{\pi n}\sin\frac{2\pi nt}{T}.$$

Fig. 9·3 shows the result of graphical addition of only the first three terms for half a period. (See also Problem 9.8).

⊙

9.3. B Fourier Transforms. If the function $y=f(t)$ is *not periodic*, then it amounts to $T\to\infty$, that is $\omega_0\to 0$. In that case the harmonic frequencies $\omega_0, 2\omega_0, 3\omega_0 \ldots n\omega_0 \ldots$ form a continuous set. Fourier showed that Eq. (9.29) can be adopted for this case also, replacing $n\omega_0$ by ω, and a_n by $F(\omega)\,d\omega$, and changing the summation to integration :

$$y=f(t)=\int_{-\infty}^{\infty}F(\omega)\,e^{i\omega t}d\omega \qquad \ldots(9{\cdot}31)$$

Eq. (9·30) then becomes

$$F(\omega)=\int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\, dt \qquad \text{...(9·32)}$$

These equations are called *Fourier transforms*. The first expresses a non-periodic function $f(t)$ in terms of an integral of harmonic time functions of various ω values whose number in a range $d\omega$ is $F(\omega)\, d\omega$. Function $F(\omega)$ is called the *density of states* at ω. Eq. (9·32) gives this density of states at frequency ω if $f(t)$ is given

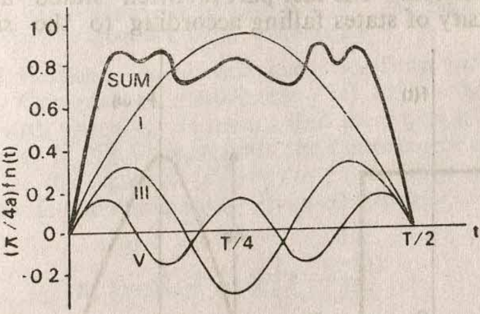


Fig. 9·3.

Plots of first three member of the Fourier series and their sum for $t=0$ to $T/2$ for Example 9·4.

You may notice the symmetry in these two equations (9·31) and (9·32). The usefulness of using complex number representation of oscillations may also be seen.

One simple example is a single square pulse of height h and width b. Taking $t=0$ at the centre of the pulse gives

$$f(t)=h \quad \text{for } -\tfrac{1}{2}b<t<\tfrac{1}{2}b$$
$$=0 \quad \text{elsewhere.}$$

Eq. (9·32) then gives

$$F(\omega)=\int_{-b/2}^{+b/2} h \exp(-i\omega t)\, dt$$

This leads to

$$F(\omega)=hb\frac{\sin\psi}{\psi}, \text{ where } \psi=\tfrac{1}{2}\omega b \qquad \text{...(9·33)}$$

The function $(\sin\psi)/(\psi)$ is called a *sinc* function. It equals 1 at $\psi=0$ and has zeros at $\psi=n\pi$, which means $\omega=(2\pi/b)\, n$. Fig.

(9·4) shows the pulse and its Fourier transform. Note that the peak of $F(\omega)$ equals the area hb of the pulse, but the width of the central maximum of $F(\omega)$ varies inversely as the width of the pulse.

Substituting for $F(\omega)$ in Eq. (9·31), and taking the corresponding $+\omega$ and $-\omega$ components together, gives

$$f(t) = \int_0^\infty 2hb \, \frac{\sin \frac{1}{2}\omega b}{\frac{1}{2}\omega b} \, d\omega \cos \omega t \qquad \ldots(9{\cdot}34)$$

Thus the pulse is the sum of an infinite number of cosine waves of frequencies ranging from 0 to ∞ and amplitudes falling according to the *sinc* function. The last part is often stated as 'amplitudes equal but density of states falling according to the *sinc* function.'

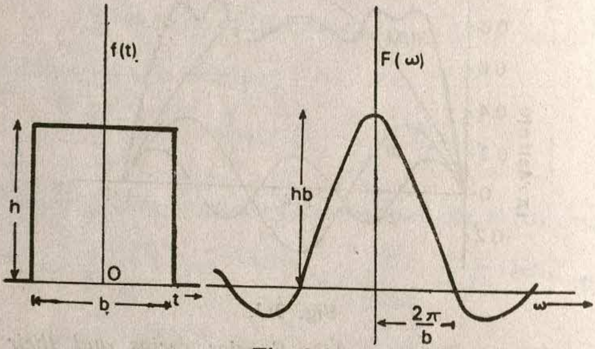


Fig. 9·4.

A square pulse (left) and its Fourier transform (right).

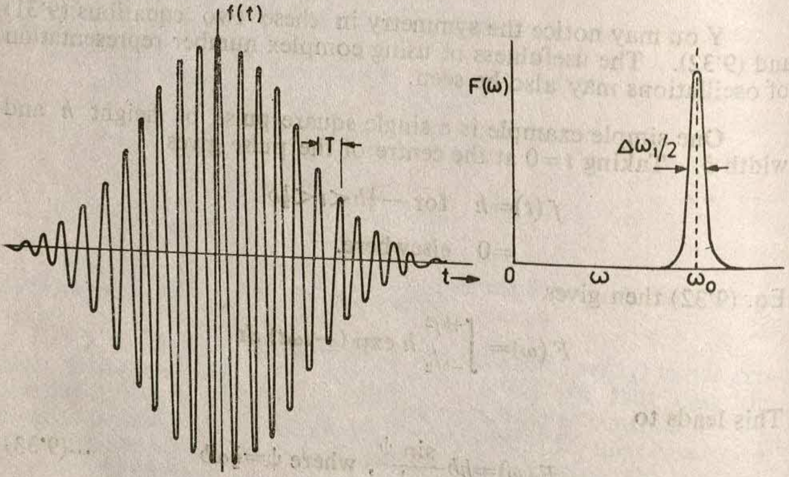


Fig. 9·5.

(left) An oscillating function $f(t)$ of period T and finite spread b, and (right) The Fourier transform with peak at ω_0 $(=2\pi/T)$ and half-intensity width $\Delta\omega_{1/2}$ proportional to $1/b$.

Of special interest is the Fourier transform of an oscillating function which tapers down at the ends, as in Fig. 9·5 (*a*). Here t could be time or distance; accordingly T could be period or wavelength. The total length (of time or distance) is b. The Fourier transform is shown in Fig. 9·5 (*b*). Its peak position ω_0 equals $2\pi/T$ (as expected). What is significant is that its spread $\triangle\omega_1/_2$ is related to b ; it is inversely proportional to b. In fact if $b_1/_2$ is the width over which $f(t)$ falls to $2^{-1/2}$ of the maximum, then $\triangle\omega_1/_2/\omega_0$ is equal to $T/b_{1/2}$.

Actually the Fourier transform of the case has also a negative ω part which is the reflection of what is shown in Fig. 9·5 (*b*). We omit it since it is understood in the complex representation of oscillations.

9·4. Wave Motion. A disturbance travelling with time in space is a wave. Consider a disturbance $f(t)$ at $x=0$, travelling along the x axis with speed v. It means that to travel a distance x it takes time $\tau=x/v$. If $f(x, t)$ represents the disturbance at x at time t, then it must be the same at $f(t-\tau)$ or $f(t+\tau)$ depending upon whether the wave is travelling away from or towards $x=0$. So in general

$$f(x, t)=f(t\mp\tau)=f\left(t\mp\frac{x}{v}\right) \qquad ...(9·35)$$

Eq. (9·35) says nothing about the nature of the disturbance ; it could be a single pulse of any shape, or a harmonic disturbance. All it says is that it travels without change of form with speed $+v$ or $-v$. As we have seen, a general $f(t)$ can be shown to be the sum of several harmonic disturbances of different ω values. If v varies with ω, then the shape of the disturbance at distant points will be different from that at the origin. Presently we ignore this.

From Eq. (9·35) we get

$$\frac{\partial^2 f}{\partial t^2}=f''\left(t-\frac{x}{v}\right);\ \frac{\partial^2 f}{\partial x^2}=\frac{1}{v^2}f''\left(t-\frac{x}{v}\right)$$

Therefore

$$\frac{\partial^2 f}{\partial t^2}=v^2\ \frac{\partial^2 f}{\partial x^2} \qquad ...(9·36)$$

This is called the *differential wave equation*, of which Eq. (9·35) is a solution.

Consider two other equations for wave motion :

$$\psi=f\left(t-\frac{lx+my+nz}{v}\right) \qquad ...(9·37)$$

and

$$\psi=\frac{A}{r}f\left(t-\frac{r}{v}\right) \qquad ...(9·38)$$

The first is a case where f at a given time is the same over a plane $lx+my+nz=r$. It shows the travel of the disturbance along normal to this plane. This satisfies Eq. (9·36) as we had seen with Eq. (8·7). The second, Eq. (9·38), is a case where f at a given time is the same over a sphere of radius r. It shows the radial travel of a spherical wave, and in this case the amplitude is a function of r. It can be shown that this also satisfies Eq. (9·36). In such cases the form of the differential wave equation (9·36) becomes

$$\frac{\partial^2\psi}{\partial t^2}=v^2\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right) \qquad \text{...(9·36 a)}$$

Let us limit ourselves to a harmonic wave, and one which travels along one line, say z-axis. If we are dealing with electric field E in an electromagnetic wave, we write

$$E=E_0\cos 2\pi\nu\left(t-\frac{z}{v}\right)\equiv E_0\cos\omega\left(t-\frac{z}{v}\right)\equiv E_0\cos\frac{2\pi}{T}\left(t-\frac{z}{v}\right) \qquad \text{...(9·39)}$$

Recall that $2\pi\nu=\omega=2\pi/T$, and all that we have done is to replace t by $t-(z/v)$ to change from harmonic *oscillation* to harmonic *wave*.

Eq. (9·39) shows that E is *periodic in time* (repeating at intervals T), and also *periodic in space*, repeating at intervals vT. This interval $vT=\lambda$ is called wavelength, and Eq. (9·39) is equivalent to

$$E=E_0\cos 2\pi\left(\nu t-\frac{z}{\lambda}\right)\equiv E_0\cos\left(\omega t-\frac{2\pi z}{\lambda}\right)\equiv E_0\cos 2\pi\left(\frac{t}{T}-\frac{z}{\lambda}\right) \qquad \text{...(9·40)}$$

Complex Representation. Eq. (9·40) may be written as

$$E=E_0\exp\left[-2\pi i\left(\nu t-\frac{z}{\lambda}\right)\right] \qquad \text{...(9·40 a)}$$

The assumption is that the complex conjugate $E_0^*\exp\left[2\pi i\left(\nu t-\frac{z}{\lambda}\right)\right]$ is to be added on the right-hand side, as in Eq. (9·9).

If is often convenient to introduce the quantity

$$k\equiv 2\pi/\lambda \qquad \text{...(9·41)}$$

which gives the harmonic wave equation as

$$E=E_0\cos(\omega t-kz) \qquad \text{...(9·42)}$$

Notice that $\nu=\lambda/T=(2\pi/k)\div(2\pi/\omega)=\omega/k$.

Phase and Wavefront. In the harmonic wave equations (9·39) or (9·40) the 'argument' of the cosine is called *phase* of the disturbance :

$$\phi\equiv 2\pi\nu\left(t-\frac{z}{v}\right)=\left(\omega t-\frac{2\pi z}{\lambda}\right)=2\pi\left(\frac{t}{T}-\frac{z}{\lambda}\right) \qquad \text{...(9·43)}$$

From this we have

$$\left(\frac{\partial \phi}{\partial t}\right)_z = 2\pi\nu = \omega = \frac{2\pi}{T} \qquad \ldots(9{\cdot}44)$$

$$\left(\frac{\partial \phi}{\partial z}\right)_t = -\frac{2\pi}{\lambda} \qquad \ldots(9{\cdot}45)$$

$$\left(\frac{\partial z}{\partial t}\right)_\phi = \nu\lambda = \lambda/T = v \qquad \ldots(9{\cdot}46)$$

These enable us to deduce the phase change $\Delta\phi$ with time (at a given place), $\Delta\phi$ with distance (at a given time), and the travel speed of a given phase. The v in Eq. (9·39) is called the *phase velocity* for this last reason; it is the same as wave velocity.

The locus of constant phase in space *at a given time* is called a *wavefront*. In waves over water surface, for example, a crest-line is one wavefront, a trough-line is another, and so on. In Eq. (9·39) or (9·40) the wavefront is thus given by z=constant, *i.e.*, a plane parallel to xy plane. In Eq. (9·37) the wavefront is $lx+my+nz$=constant, although the wave is not harmonic. It is a plane whose normal has direction cosines l, m, n. In Eq. (9·38) the wavefront is r=constant, which is a sphere.

There are situations where the amplitude over a wavefront may not remain constant. Yet constancy of phase defines a wavefront. A wavefront partly transmitted and partly reflected at a boundary is an example.

Example 9·5. Blue Light shows $\lambda=5{\cdot}0\times10^{-5}$ cm in vacuum. It enters a medium where its speed is $2{\cdot}0\times10^{8}$ m/s. Calculate the quantities (i) λ, (ii) ν (iii) ω, (iv) T, (v) $\Delta\phi$ for time gap $1{\cdot}2\times10^{-14}$ s, (vi) $\Delta\phi$ for spatial distance $4{\cdot}0\times10^{-6}$ cm.

Solution. (i) λ in the second medium$=5{\cdot}0\times10^{-5}\times\dfrac{2{\cdot}0}{3{\cdot}0}$cm

$=3{\cdot}3\times10^{-5}$ cm

(ii) $\nu=\dfrac{c_1}{\lambda_1}=\dfrac{c_2}{\lambda_2}=\dfrac{3{\cdot}0\times10^{8}ms^{-1}}{10^{-7}m}=6{\cdot}0\times10^{14}s^{-1}$

(iii) $\omega=2\pi\nu=3{\cdot}8\times10^{15}s^{-1}$

(iv) $T=1/\nu=1{\cdot}7\times10^{-15}s$

(v) $\Delta\phi=\omega\Delta t=0{\cdot}45$ radian

(vi) $\Delta\phi=-(2\pi/\lambda)\,\Delta z=-5{\cdot}0$ radian. ⊙

9·5. Waves from Atoms and Molecules. Light shows the same behaviour as electromagnetic radiations sent out by electronic oscillator circuits—same as regards speed of propagation, reflection, refraction and interaction with matter. Some differences arise because of large difference in frequency—electronic circuits in the highest frequency range have $\nu \sim 10^{10}\ s^{-1}$, while in middle range of the visible light $\nu \sim 5\times10^{14}\ s^{-1}$. We shall, however, refer to a more fundamental difference between the two cases.

In an electronic oscillator antenna the swarm of free electrons vibrate *in unison* and in a specified direction (along the length of the antenna). So if the wave due to one oscillating electron is E_{∞} cos $(\omega t-kz)$, that due to N electrons is just NE_{∞} cos $(\omega t-kz)$.

In the case of light, which is emitted by atoms and molecules, the situation differs in two respects. Firstly, the waves from separate atoms produces E fields in different orientations in the plane normal to the line of sight; *i.e.*, there is *randomness in the orientation* of the atomic dipole oscillators. Secondly, the waves sent by separate atoms are *random in phases*. This double randomness results in the radiations becoming isotropic in space (unlike Fig. 8·5), and also unpolarized. Even more ticklish point is that an atom does not go on producing a wave for long, and each time it sends a new wave it will have random phase relation with the previous one. A typical wave may last for, say, 10^{-9} s or so, which means that it may have a spread in space $\sim$30 cm or so.

If we can somehow orient the atomic dipole oscillators, as by applying a magnetic field (Zeemann effect) the emitted radiation will have a specific polarization, *i.e.*, a chosen direction of E in the plane normal to line of sight. And if we can arrange that atoms do not emit light at their own sweet will but do so under an external agency (as in lasers) then the emissions by different atoms may produce waves in phase agreement. These, however, are special situations.

9·6. Wave Groups and Wave Packets. An oscillation represented by $E=A\cos 2\pi\nu t$ is an idealised case of an oscillation going on from time $t=-\infty$ to $t=+\infty$ without change in amplitude. The corresponding wave represented by

$$E=A\cos 2\pi\left(\frac{t}{T}-\frac{x}{\lambda}\right) \qquad \ldots(9{\cdot}47)$$

is also therefore an idealised case of E repeating in space from $x=-\infty$ to $x=+\infty$ with periodicity λ and constant amplitude A. Fig. 9·6 (a) represents such an idealised wavetrain.

In practice no oscillation has an infinite extension in time and hence no wavetrain has infinite extension in space. In Fig. 9·5 we treated a practical oscillation. Here Fig. 9·6 (b) and 9·6 (c) represent wavetrains of finite length if the horizontal axis is considered as distance x—these are known as *wave groups or wave packets*. Fig. 9·6 (d) gives a wavetrain of very short length, showing a little more than one oscillation. This is known as a *pulse*.

It is to be noted that long or short wave groups are due, respectively, to small or large damping in the oscillator at the source. This character of the oscillator is represented by *quality factor* (Q factor), treated in § 9·2. For the wave of Fig 9·6 (*a*), we have $Q=\infty$; for the wave of Fig. 9·6 (*b*), the value of Q is about 10^2. For radiations given out by atoms and molecules typical Q factors are of the order of 10^6 to 10^{10}.

Now consider a wave group in which wavelength is λ and the amplitude falls to negligible fraction* of the maximum in a length

---

* We could set any value for this fraction. If it is set at $e^{-1/2}$, then m becomes identical with the quality factor Q.

m times λ from the centre, so that m gives a sort of measure for the length of the wave group. Then, as stated with Fig. 9·5, the wavegroup may be treated as the superposition of an infinitely long wavetrain of wavelength λ with infinitely long wavetrains of wave lengths spread on both sides of λ with successively diminishing amplitudes. Fourier transform gives the amplitude distribution.

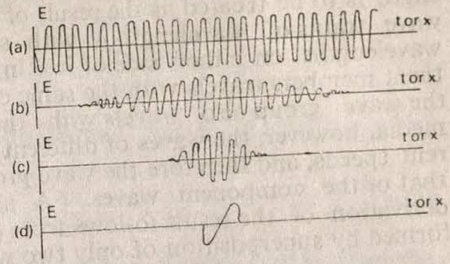


Fig. 9·6.

Wavetrains of different lengths : (*a*) *an ideal wave*, (*b*) *a wave group*, (*c*) *a shorter wave group, and* (*d*) *a pulse*

If the wavegroup is short, the spread of wavelengths to be included in this representation is larger. We then say that the wavelength of this wave is not sharp ; correspondingly the frequency is not sharp (*cf.* Fig. 9·5). On the other hand, if the wave group is a long one (say $10^{10}\lambda$ long), we say that the radiation has a sharp frequency.

Let us put it conversely. A wave with sharp range of frequencies means a long wavetrain ; one with wider range of frequencies means a shorter wavetrain ; and one with very wide range of frequencies means a pulse. An example in the last class is white light. Here we have a total spread (violet to red)$\sim$4000 A to $\sim$8000 A. So the wavetrain would be a pulse. In contrast, for most atomic spectral lines the half-intensity width is$\sim 10^{-2}$ A^0, and some are as narrow as 10^{-5} $A°$. Since $\lambda \sim 5000$ $A°$, we get $Q \sim 3\times10^6$ to 3×10^9. The wavetrains for them have length $\sim Q\lambda$.

A mathematically useful way of describing a wave packet (or wave group) is the Fourier transform :

$$E(x,t)=\int_{-\infty}^{+\infty} A(k)\, e^{\,i(kx-\omega t)}\,dk \qquad \ldots(9{\cdot}48)$$

where $k=2\pi/\lambda$ and $\omega=2\pi/T$. (This is similar to Eq. 9·31 as applied in Fig. 9·5). Here $A(k)$ is a function of k (hence of λ), having the maximum value at a particular $k=k_0$ and falling off on either side.

A simple way to see the above is to consider the superposition of two waves of equal amplitude with wavelengths λ and $\lambda+\Delta\lambda$. They get out of phase by π in a length L given by

$$\frac{L}{\lambda}-\frac{L}{\lambda+\Delta\lambda}=\tfrac{1}{2} \;\Rightarrow\; \frac{\lambda}{\Delta\lambda}=\frac{2L}{\lambda}=\frac{b}{\lambda}$$

Thus the total spread b is inversely proportional to $\Delta\lambda$, also the number of wavelengths in the group (b/λ) equals to ratio $\lambda \mid \Delta\lambda$. Of course, with just two waves taken, the 'group' will repeat again and again (these are *beats*), hence an infinity of sinusoidal wavetrains as prescribed by the Fourier transform are needed to produce a group (which will not repeat).

9·7. **Wave Velocity and Group Velocity.** A wave group of finite ength is to be treated as the result of superposition of infinitely long wave trains of wavelength λ with similar waves of neighbouring wavelengths on either side of λ. In vacuum, each of these constituent members travels with the same common speed c, and therefore the wave group also travels with the same speed c. In material media, however, the waves of different wavelengths travel with different speeds, and therefore the wave group has a speed different from that of the component waves. A simple (though, not rigorous) derivation of the result follows from consideration of a wave group formed by superposition of only two waves.

In Fig. 9·7, on the left side top we show two waves of wavelengths λ and $\lambda+\triangle\lambda$ at instant $t=0$. One pair of crests of these waves coincide at position P, and on either side of P the crests get more and more out of step. The coinciding crests are marked a_1, a_2 and the next crests b_1, b_2 are seen out-of step by $\triangle\lambda$. The wave group, shown at left-side bottom, has its centre at P.

Now let the component waves of wavelengths λ and $\lambda+\triangle\lambda$ travel to the right with speed u and $u+\triangle u$ respectively. In material media $\triangle u/\triangle\lambda$ is positive, *i.e.*, waves of larger wavelength travel faster. At the instant $t=0$, the crests b_1 and b_2 just behind P had a gap $\triangle\lambda$. As the waves travel forward, this gap covers up at a rate $\triangle u$ per second, so that after time $t_0=\triangle\lambda/\triangle u$ the crests b_1 and b_2 will coincide. Fig. 9·7 (top right) shows the situation at this particular time. Now crests a_1, a_2 have a gap of $\triangle\lambda$. Henee the maximum amplitude of the resultant *group* is not at P', but is at Q', as shown at the bottom right in Fig. 9·7. The centre of the group (Q') is one wavelength (λ) behind P'.

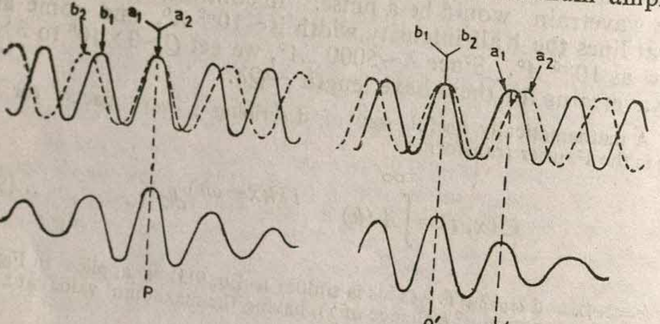


Fig. 9·7.

Group and waves have different speeds

The individual *wave* of wavelength λ has travelled a distance

$$PP'=ut_0$$

In the same time the *group* has travelled a distance

$$PQ'=PP'-Q'P'=ut_0-\lambda$$

$$\therefore \quad \text{group velocity}=\frac{\text{distance}}{\text{time}}=u-\frac{\lambda}{t_0}$$

Calling group velocity c_g and substituting $t_0=\Delta\lambda/\Delta u$, which in the limit is $d\lambda/du$, we get

$$c_g=u-\lambda\frac{du}{d\lambda} \qquad \text{...(9·49)}$$

Since $du/d\lambda$ is generally positive, the *group* travels slower than the individual *waves*. In more dispersive media, $du/d\lambda$ is large, and hence the difference $u-c_g$ may be quite large. In vacuum $du/d\lambda=0$, and hence $c_g=u=c$.

Two related results are of interest. Since $\omega=uk$, one can show that

$$c_g=d\omega/dk \qquad \text{...(9·50)}$$

Also, since $u=c/n$ (where n is refractive index), one gets from Eq.(9·49)

$$c_g=u\left(1+\frac{\lambda}{n}\frac{dn}{d\lambda}\right)=u\left(1+\frac{\lambda_0}{n}\frac{dn}{d\lambda_0}\right) \qquad \text{...(9·51)}$$

where λ_0 refers to the wavelength in vacuum.

It is to be noted that in all experiments on direct measurement of speed of waves it is the *group velocity* c_g which is measured. This is the speed with which energy is transferred in the medium. In contrast, the refractive index equation $n=c/u$ involves *wave velocity* (or phase velocity) u, as we shall see in the next chapter.

Example 9.8. For EDF glass the refractive index data are as follows :

λ (yellow)$=5873$ A° $\quad n=1·7003$

λ (blue) $=4861$ A° $\quad n=1·7168$

Deduce c_g-u in the neighbourhood of $\lambda\sim5000$ A°, and compare this with the change of u itself from yellow to blue.

Solution. From eq. (9·51),

$$\frac{c_g-u}{u}=\frac{\lambda}{n}\frac{dn}{d\lambda}$$

From the data, for order-of-magnitude purpose, we get

$$n\,(5000\ A^\circ)=1.71,\ \frac{dn}{d\lambda}\approx-\frac{·0165}{1000\times10^{-8}}\approx-1600$$

$$\therefore \quad \frac{c_g-u}{u}=-\frac{5000\times10^{-8}}{1·71}\times1600\approx-5\times10^{-2}$$

The difference from u (yellow) to u (blue) we write Δu. Then

$$\frac{\Delta u}{u}=-\frac{\Delta n}{n}=\frac{0·165}{1·7}\approx-1\times10^{-2}=\frac{1}{5}\frac{c_g-u}{u}$$

⊙

9·8. Direct Measurement of Wave Velocity. The methods described in Chapter 7 measured the group velocity, since speed of energy transfer was essentially measured. The wave or phase velocity is given by $\nu\lambda$, where ν is the frequency and λ the wavelength.

The first accurate experiment to determine wave velocity *directly* was done by Essen and Gordon-Smith in 1947 using radio-waves of frequency $\sim$300 *MHz* from a frequency standard and arranging a resonant cavity for which λ was known from the dimensions. The accuracy was 1 part in 10^6.

Froome in 1958 used a standing wave pattern to measure λ. Fig. 9·8 shows his arrangement. The oscillator was of frequency $\sim$100 *GHz* (10^{11} *Hz*) and waves emitted from the horns *H*,*H* form standing waves. The double horn detector *A* moving along the scale measures the variations in the signal and hence determines λ.

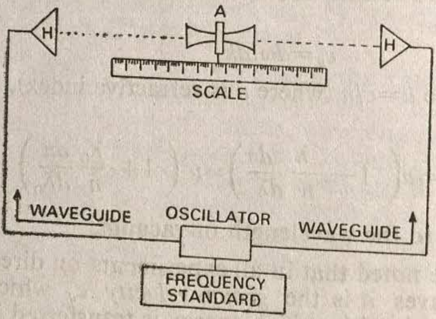


Fig. 9·8.

Determination of Phase Velocity.

9·9. Necessity for Representation of Waves by Complex Quantities. From § 9·1 *A*, § 9·2 *A* and § 9·3 *A* we find that complex representation of waves is a matter of great mathematical *convenience*. When we switch to Fourier transforms (§ 9.3 *B*), we find that in analysing a non-periodic function the only way we can proceed is through complex representation of $y=f(t)$ and getting the analysed density of states $F(\omega)$ at frequency ω as a complex representation. Yet, after this mathematical exercise is over, we may still change the results to the usual real-wave forms. It means that complex representation is merely a mathematical tool of great *convenience*, but the oscillations are real.

However, in quantum mechanics we face a different situation. Unlike the equations

$$y=a\cos(\omega t-kx), \text{ or } E=E_0\cos(\omega t-kx)$$

where y may be a displacement and E an electiric field, in quantum mechanics we have parallel expressions like

$$\psi=(a+ib)\cos i(\omega t-kx) \qquad (9{\cdot}50)$$

where ψ itself has no direct real-physics meaning, but the quantity $|\psi^*\psi|$ represents *probability density* of occurrence of a certain event. It is not that there is some real-wave representation for which we write eq. (9·50) for convenience; here eq. (9·50) is itself the original expression to represent the wave. This representation comes as a *necessity*, not as just a mathematical convenience.

What does ψ in eq. (9·50) represent? Not displacement of anything ; not electrical or magnetic or any field; not even probability. In fact, since ψ is complex, it cannot represent any real quantity. Why use such an expression at all, then ? The answer is that nature does not permit us to give description of the behaviour of particles and radiations at the micro-level in simple terms directly. Yet if a description in the form of eq. (9·50) is used then $|\psi^*\psi|$ gives us information related to observable real physics. We will come to that in Chapter 23. In the meantime it is good to appreciate that complex representation, devised as a pure mathematical tool, has become a *necessity* in describing atomic and nuclear phenomena in quantitaive manner.

PROBLEMS

9·1. For the electrical circuit of Fig. 9·1, the absence of R is an ideal not attainable in practice. Explain this. What relevance has this statement with the fact that no atomic radiation has zero half-width.

9·2. (*a*) Express the oscillation $\xi=5\cos(1000t+0{\cdot}35)$ in the forms $A_c\cos\omega t+A_8\sin\omega t$ and $a\exp i\omega t$.

(*b*) Express the oscillation $\xi=(4+3i)\cos i\,500\,t$ in the forms $\xi=A\cos(\omega t+\phi)$ and $\xi=A_c\cos\omega t+A_s\sin\omega t$.

9·3. An oscillation is described by $\xi=(50+30i)\exp(400+5i)t$ Deduce the initial amplitude and phase, relaxation time, and Q value of the oscillation.

9·4. The electrical field in a plane e.m. wave is given by $E=4{\cdot}0\times10^{-2}\times e^{-{\cdot}005t}\cos(3{\cdot}0\times10^{14}t)$. If it is expressed as $E=a\cos i(\omega+ik)t$, deduce a, ω and k. Also deduce $|E^2|$, and the intensity at $t=0$, if E is given in volt/m.

9·5. Prove the validity of Eq. (9·28 *b*) if $f(t)$ is given by Eq. (9·27).

9·6. (*a*) In the complex representation, Fourier series involves summation over n from $-\infty$ to $+\infty$. Explain the inclusion of negative n values.

(*b*) The Fourier transform shown in Fig. 9·5 (*b*) has a counterpart at $-\omega_0$ also. Explain.

9·7. If the function $f(t)$ of Fig. 9·5 (*a*) were to repea
intervals b, then it represents beats, which arise from the summ

of two harmonic (ideal) waves. If it does not repeat, then it arises from a summation of harmonic waves with density of states shown by $F(\omega)$. Examine critically.

9·8. The periodic oscillation of Example 9·4 (p.111) has the Fourier expansion $f(t)=\Sigma\,(4a/\pi n)\sin(2\pi nt/T)$, summation over odd n only. Show that $f(T/2-t)=f(t)$, and $f(T/2+t)=-f(t)$. Also, for $t=T/60$, deduce the sum of the terms falling in the ranges $(2\pi nt/T)=0$ to 90°, 90° to 270°, and 270° to 450°. Comment on the results.

9·9. A wave is represented by $\psi=3{\cdot}0\times10^{-3}\cos(8{\cdot}4\times10^{13}t+2{\cdot}8\times10^{5}z)$ volt/m. Compute the (*i*) amplitude, (*ii*) frequency, (*iii*) wavelength, and (*iv*) wave velocity (assuming t given in sec and z in metre).

9·10. Define a wavefront. Cite an example to show that it is not necessary that amplitude over a wavefront may be constant.

9·11 In what respects do atomic oscillators differ from an electronic oscillator ?

9·12. For a certain material the refractive index n varies in the visible range with $dn/d\lambda=1{\cdot}2\times10^{5}\,m^{-1}$ and mean $n=1{\cdot}532$. Deduce (i) the mean wave velocity t', (ii) the mean group velocity c_g', (iii) the difference $\triangle c'$ for change of λ from 5000 to 6000 A, and (iv) the ratio of $c_g'-c'$ to $\triangle c'$.

9·13. Deduce the expressions for group velocity when the phase velocity u is related to λ through

(*i*) $u=a\lambda^{1/2}$ (surface waves controlled by gravity)
(*ii*) $u=a\lambda^{-1/2}$ (surface waves controlled by surface tension)
(*iii*) $u=a+b\lambda$ (a hypothetical case)

9·14. If $\omega=2\pi/T$ and $k=2\pi/\lambda$, show that

$$u=\omega/k \text{ and } c_g=d\omega/dk$$

10

Reflection and Refraction of Waves

Most of the material in this chapter will be applicable to all kinds of waves—longitudinal or transverse, and mechanical or electromagnetic. The word '*disturbance*' will therefore be used here for change in the relevant vector property of the medium; it could be the electric field E in electromagnetic waves or particle displacement in an elastic wave.

The propagation of a wave in a homogeneous and isotropic medium is simple. The next problem is to understand reflection and refraction at the boundary of different media. In this chapter we shall not consider anisotropic media ; also the wavelength will be treated as very large compared with the atomic spacings in the media, so that the media are taken as continuous isotropic media. Only a broad *macroscopic* understanding will thus be developed, using a principle first given by Huygens.

10·1. Huygens' Principle. A wave-front is *the locus of points in space having the same phase* of a given periodic variation. Huygens gave a simple principle which really is a *geometrical procedure* for determining the progress of a wave-front in a given medium and on reflection and refraction at the separation of two media.

Huygens' principle may be stated in three parts :

(*a*) *Every point on a given wave-front may be regarded as the origin of a new disturbance.*

(*b*) *The new disturbance from each point spreads out in all directions with the speed proper for the medium ; these disturbances are called secondary wavelets.*

(*c*) *The new wave-front at any later time is given by the forward envelope of the secondary wavelets.*

In Fig. 10·1 AB is section of a wave-front at any instant t. To find the wave-front at an instant $t+\Delta t$, the procedure is as follows :

(*a*) At several points on AB draw circles of radii $v\Delta t$, to represent sections of the secondary spherical wavelets, where v is speed of the wave in the given medium.

(*b*) Draw the common tangents *CD*, *C′D′*; they represent sections of 'envelopes' of the wavelets.

(*c*) If the source is towards the left. *CD* is the 'forward' envelope and gives the new wave-front. If the source is towards the right, *C′D′* is the new wave-front.

Huygens' principle does not explain why the 'forward' envelope alone is to be taken, and back wave is to be neglected. Further, in taking the envelope we in effect neglect all parts of the secondary wavelets except those which fall on the envelope. Thirdly, if the wave-front is limited in width we do not know what to do to the envelope at the ends.

A more mathematical form of the principle was later given by Kirchoff to account for the absence of the back-wave, but we shall not go into it here. Fresnel's procedure dealt with the last problem mathematically ; we shall study it under the title 'diffraction' of light, which occurs when wavefronts are limited in width. For the present we shall see what the wave theory with Huygens' principle can explain.

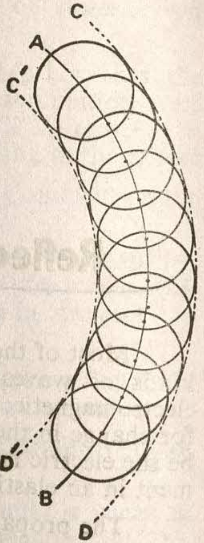


Fig. 10·1. *Illustrating Huygens' Principle*

10·2. Laws of Reflection from Wave Theory. The familiar laws of reflection are as follows :

(i) The reflected ray is in the plane defined by the incident ray and the normal to the surface* at the point of incidence.

(ii) The angle of reflection is equal to the angle of incidence.

In Fig. 10·2, W_1W_2 represents section of a plane wave-front whose normal W_1A lies in the plane of paper. The wave-front is advancing towards a plane surface S_1S_2, whose normal *AN* also lies in the plane of paper. The angle W_1AN is the angle of incidence *i*; the plane of paper (containing W_1A and *AN*) is the plane of incidence.

The advancing wave-front strikes different point on the surface S_1S_2 at different times. Consider the moment when a point *B* is just disturbed by the wave-front. At this moment the secondary wavelet from *B* has zero size. Points to the left of *B* have been disturbed earlier ; if *A* is a point at distance *x* from *B*, then the time interval *t* by which *A* was disturbed earlier than *B* is given by

---

* It may be any surface separating one medium and another. It need not be a 'mirror'.

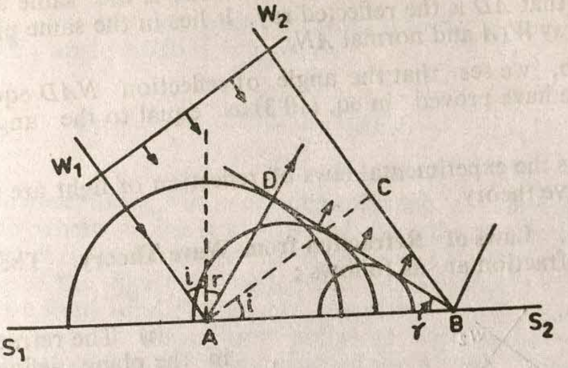


Fig. 10·2.

Reflection of a Wave

$$t=\frac{W_2B-W_1A}{v}=\frac{BC}{v}=\frac{x \sin i}{v} \qquad \ldots(10{\cdot}1)$$

where AC is drawn perpendicular from A on W_2B and v is the speed of wave in the medium.

The secondary wavelet from A grows in this time to a sphere of radius AD given by

$$AD=t.v=x \sin i. \qquad \ldots(10{\cdot}2)$$

Now from B draw a tangent BD to the circular section, which touches the circle at D. The angle ABD is given by

$$\sin \angle ABD=\frac{AD}{AB}=\frac{x \sin i}{x}=\sin i \qquad \ldots(10{\cdot}3)$$

We note that $\angle ABD$ does not depend upon the distance x; hence we could imagine secondary wavelets at an infinite number of points along the line BA, and the line BD would be tangent to the trace of *all* those wavelets. BD is thus a *common* tangent (envelope).

Further, if the figure is moved towards normal to the plane of paper, BD would describe a plane, which would touch all the wavelets arising from the surface S_1S_2. Therefore, the plane described by BD is the reflected wave-front. As time further advances, this wave-front would travel along its normal, *i.e.*, along AD, because each of the spherical wavelets would grow and eq. (10·3) shows that angle ABD is constant ($=i$).

Noting that normal to a wave-front is the same as ray, we conclude that AD is the reflected ray. It lies in the same plane as the incident ray W_1A and normal AN.

Also, we see that the angle of reflection NAD equals ABD, which we have proved in eq. (10·3) as equal to the angle of incidence.

Thus the experimental laws of reflection of light are supported by the wave theory.

10·3. Laws of Refraction from Wave Theory. The familiar laws of refraction are as follows :

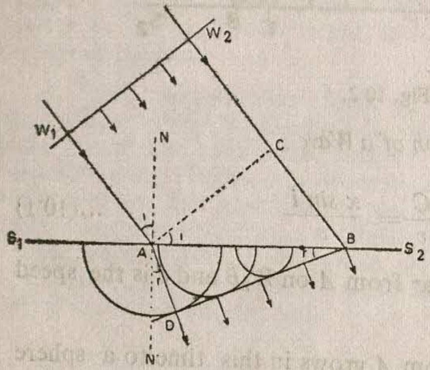


Fig. 10·3
Refraction of a *Wave*

(i) The refracted ray is in the plane defined by the incident ray and the normal to the surface at the point of incidence.

(ii) The sine of angle of incidence bears a constant ratio to the sine of the angle of refraction (Snell's Law).*

In fig. 10·3, S_1S_2 is section of a plane surface, whose normal AN lies in the plane of paper. This surface separates two media, the speed of waves in the upper medium being v and that in the lower medium v'. W_1W_2 represents section of a plane wave-front whose normal W_1A lies in the plane of paper ; the angle of incidence is i.

The wave-front W_1W_2 strikes different points of S_1S_2 in succession. Consider the instant at which the point B is just disturbed. At this instant the secondary wavelet at B has zero size. If A is a point such that $AB=x$, then the time interval t by which A was disturbed earlier than B is given by

$$t=\frac{W_2B-W_1A}{v}=\frac{CB}{v}=\frac{x \sin i,}{v} \qquad \text{...(10·4)}$$

where AC is drawn perpendicular from A on W_2B.

In this time t the secondary wavelet from A *in the second medium* grows to a size AD given by

$$AD=tv'=(x \sin i)\frac{v'}{v}$$

---

* The law, first given for light, is found true for all waves.

We now draw a tangent *BD* from *B* to the section of wavelet at *A*. The angle $ABD = r$ is given by

$$\sin r = \frac{AD}{x} = \frac{x \sin i}{x} \frac{v'}{v} = (\sin i) \frac{v'}{v} \qquad \ldots(10{\cdot}5)$$

We note that $\angle r$ does not depend upon x. Therefore, we may take *A* anywhere to the left of *B* and draw the wavelets at the instant when *B* is just disturbed ; in every case *BD* will be the common tangent. If the figure is now moved perpendicular to the plane of paper, we shall find *BD* describing a plane which touches all the secondary wavelets arising from points on surface S_1S_2—it thus describes the envelope, *i.e.*, the refracted wave-front. The normal to *BD* gives the ray *AD* and it lies in the plane of paper. This proves the first law of refraction.

Further, $r = \angle N'AD$ = angle of refraction. Hence, eq. (10·5) gives

$$\frac{\sin i}{\sin r} = \frac{v}{v'} = \text{constant } (n) \qquad \ldots(10{\cdot}6)$$

This proves the second law of refraction (Snell's law). It further tell us that $n = v/v'$.

In the case of light it is customary to define the absolute index of refraction by taking the case where i and r refer to vacuum and the medium concerned respectively. The speeds are denoted by symbols c and c'. Hence we put eq. (10·6) as

$$n = \frac{c}{c'} \text{ or } c' = \frac{c}{n} \qquad \ldots(10{\cdot}6\text{ a})$$

This result was a *prediction* of wave theory. Newtons' corpuscular theory gave $c' = c.n$.

Example 10·1. Sodium yellow light has wavelength 5893 A° in vacuum. This light enters a medium of $n_D = 1{\cdot}52$. Deduce the speed, wavelength and frequency of this light in the medium.

Solution. (*i*) $c' = \dfrac{c}{n} = \dfrac{3.00 \times 10^8}{1{\cdot}52} = 1{\cdot}97 \times 10^8$ m/sec

(*ii*) $\lambda' = \dfrac{c'}{\nu} = \dfrac{c'}{c/\lambda} = \dfrac{\lambda}{n} = \dfrac{5893}{1{\cdot}52} = 3877\ A^0$

(*iii*) $\nu' = \nu = \dfrac{c}{\lambda} = \dfrac{c'}{\lambda'} = 5{\cdot}10 \times 10^{14}\ \text{sec}^{-1}$ (check) ⊙

10·4. Image Formation on Wave Theory. Once we have established that laws of reflection and refraction are valid on the wave theory if wave-normals and rays are taken as equivalent, all the rules of image formation deduced from those laws become automatically valid on wave theory.

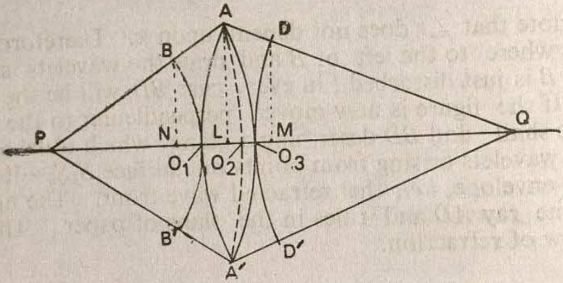


Fig. 10·4.

Passage of a wave- front through a convex lens

However, it is interesting to follow the *changes in a wave-front* in at least one typical case. In Fig. 10·4 we show this for the case of passage of a wave across a convergent lens. AO_1A' and AO_3A' are the two faces of the lens and P is a source of waves. The spherical waves spreading out from P strike the middle part of the lens earlier and outer parts later. BO_1B' represents the wave-front when it has just touched the middle part of the lens. By the time the outer ends A, A' of the lens are reached by the wave-front, the secondary wavelet at O_1 grows into size O_1O_2 given by

$$O_1O_2=\frac{BA}{c}c', \qquad \ldots(10·7\text{ a})$$

where c' is speed of light in the lens material. The wave-front at this stage is AO_2A'. Now the outer ends of the wave-front emerge out earlier than the middle parts. Hence, by the time the middle part travels distance O_2O_3, the outer parts travel a distance AD given by

$$AD=\frac{O_2O_3.}{c'}c \qquad \ldots(10·7\text{ b})$$

The emergent wave-front is thus DO_3D'. Combining eqs. (10·7 a) and (10·7 b) we see that

$$BA+AD=(O_1O_3)\frac{c}{c'}=n(O_1O_3). \qquad \ldots(10·8)$$

where n is refractive index of the lens material.

This condition combined with geometry and *first order approximations* gives the well known result

$$\frac{1}{v}-\frac{1}{u}=(n-1)\left(\frac{1}{r_1}-\frac{1}{r_2}\right) \qquad ...(10\cdot9)$$

To prove this we write the left side of eq. (10·8) as equal to NM*, which is $NO_1+O_1L+LO_3+O_3M$. The right side of eq. (10·8) is $n\,(O_1L+LO_3)$. Hence we get from eq. (10·8),

$$NO_1+O_3M=(n-1)\,(O_1L+LO_3) \qquad ...(10\cdot10)$$

If we put $h=BN=AL=DM$ to first aproximation, then with usual notation and sign convention the four lengths in eq. (10·10) are $-h^2/2u$, $h^2/2v$, $h^2/2r_1$ and $-h^2/2r_2$. Simple substitution in eq. (10·10) gives the result (10·9).

Strictly speaking, the wave-fronts AO_2A' and DO_3D' will *not* be speherical, Only their central part will fit well into a spherical shape, outer parts will not. Therefore, the emergent wave-front DO_3D', as it travels forward, would not shrink to any sharp point like Q. This statement corresponds to what is called *spherical aberration* discussed in Chapter 5 on the ray optics method.

The Principle of Equal Optical Paths. It will be noted that the aforesaid image deduction is equivalent to setting the condition that the times taken by the disturbance to reach from P to Q through different paths are the same. If we define *optical path* corresponding to any distance as refractive index n times the distance, we could write the result as

$$\int_P^Q n\,ds=\text{constant} \qquad ...(10\cdot11)$$

over all possible paths between a source point P and image point Q. This is the principle of equal optical paths from object point to image point.

The significance of *equal optical paths* is not clear in ray optics. But in the wave picture the significance is that the disturbances reaching Q through different paths all reach in same phase and therefore add up to produce a large disturbance there. This significance is in fact already included in the consideration of a *wave-front* travelling forth and converging at a point.

10·5. Simultaneous Reflection and Refraction at the Boundary of Two Media. When a wave is incident on the boundary, the Huygens secondary wavelets grow from these points *in both the media*, the growth rates depending on the speeds of the waves in the media concerned.

Now, what fraction of the incident energy is reflected and what fraction is refracted ? On what property or properties of the two media does this depend ? Further, if the incident wave is polarized with electric vector in the plane of incident ($E_{\parallel}$) or perpendicular ($E_{\perp}$) to it, then will the reflectivity be different for the two cases ? If so, what are the relations ?

---

* This assumes that lens aperture is much smaller than magnitudes of u and v.

These questions were answered in detail by Fresnel, who considered elastic waves in 'ether'. The same results, interestingly, are applicable in the electromagnetic theory also. We may state his results as follows (without proof) :

(*a*) At normal incidence, waves of both polarizations ($E_{||}$) and $E_{\perp}$ are reflected equally with

$$\frac{E\text{ (reflected)}}{E\text{ (incident)}}=\frac{n-1}{n+1} \qquad \ldots(10{\cdot}12)$$

(b) As the angle of incidence is increased, the reflectivity for the wave of polarization $E_{\perp}$ increases continuously, slowly at first and rapidly at higher i. The reflectivity depends on i and n only.

(*c*) For the wave of polarization $E_{||}$ the reflectivity at first falls with increasing i, becomes zero when i equal $\tan^{-1}n$ and then increases rapidly.

Fresnel in fact deduced precise expressions for reflectivity in cases (*b*) and (*c*) also. But their discussion will be beyond our scope.

Example 10·2. Compute the percentage reflection when a beam of light falls normally from air over a (*i*) glass surface, (*ii*) water surface.

Solution. From eq. (10·12), we get

$$\%\text{ reflection}=100\left(\frac{n-1}{n+1}\right)^2$$

Using $n=1{\cdot}50$ and $1{\cdot}33$ for the two cases we get the results 4·0% and 2·0% respectively (check). ⊙

Example 10·3. A telescope or microscope involves at least 3 lenses between the object and eye. Consider n=1·50 and deduce the minimum loss of light when lens surfaces are ideally clean.

Solution. 3 lenses have 6 faces in all. For normal incidence each face *transmits* 96% (see Ex. 10·2). Hence for 6 faces :

$$\%\text{ transmission}=100\times\left(\frac{96}{100}\right)^6=78\text{ (check)}$$

Thus the minimum reflection loss is 22%* ⊙

Fresnel in fact deduced precise expressions for reflectivity in cases (*b*) and (*c*) also. But their discussion will be beyond our scope.

10·6. Shock Waves : Cerenkov Radiation. Consider a particle

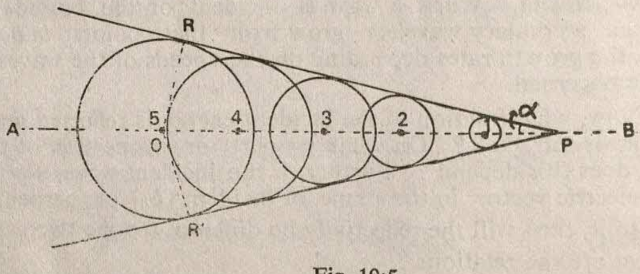


Fig. 10·5.

Development of a Shock Wave or Cerenkov Radiation moving in a medium with speed v_s which is greater than the speed v of waves in the medium.† Fig. 10·5 shows AB as the direction of the moving particle. At the instant the particle reaches P, the wavelet from O grows to size OR given by

---

* See also *non-reflecting coatings* in § 12·3.

$$OR=\frac{PO}{v_s}v$$

Since $v_s>v$, we have $OR<PO$. If PR is the tangent from P to the wavelet around O, the angle $\alpha=OPR$ is given by

$$\sin\alpha=\frac{OR}{PO}=\frac{v}{v_s} \qquad \ldots(10{\cdot}13)$$

This is independent of distance PO. Hence PR is a common tangent to all the wavelets from points along PO. Rotating the figure about line AB makes PR execute a cone with AB as axis and α as its semi-angle. The cone is a wavefront spreading out normal to its surface with speed v.

Waves as described here are called *Shock waves* since they are produced in air when bullets or bombs travel faster than speed of sound. The corresponding case in electromagetic waves is called the *Cerenkov radiation* after its discoverer It is produced when charged particles travel in a medium with speeds than the speed of light *in that medium.* In fact the cone angle α is used to determine the speed of the particle. (See Eq. 10·13).

Notice that for v_s less than v, no wavefront is produced by a moving particle.

10·7. Phase Change at Reflection. Does the reflected wave start from the boundary in the same phase as the incident wave or with some phase difference ? There are at least two different ways of seeking an answer. Fig. 10·6 shows a direct experimental method with pules sent on a string tied with another string of larger or

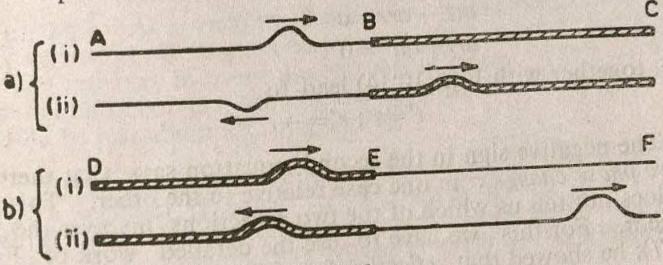


Fig. 10·6.

(a) Pulse incident from the lighter string. (b) Pulse incident from the heavier string. In each case (i) is before the pulse reaches the junction (ii) is after

---

†Law of relativity is *not* violated because v_s is presumed$>v$ in the meidum and *not* greater than c, the speed of light in vacuum.

smaller mass per unit length. The pulse travels upto the junction of the strings, and then it partly passes on to the second string and is partly reflected back on the original string. We find *experimentally* that if speed of waves in the second string is *smaller* than in the first ($v_2 < v_1$) the reflected pulse comes out with phase *reversed* with respect to the incident pulse. For $v_2 > v_1$, the reflected pulse comes out without change of phase. The corresponding *experiment* in the case of waves of light will be discussed in chapter 12.

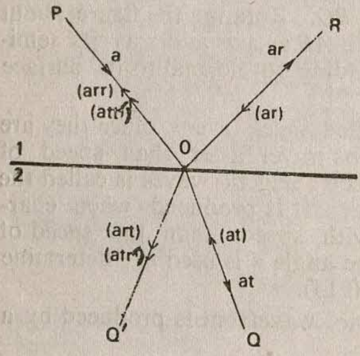


Fig. 10·7.
Showing Stokes' treatment

The second method is to use the logic first given by Stokes. In Fig. 10·7 consider a wave-normal *PO*, with wave of amplitude *a* incident on the boundary of two media 1 and 2. Let the coefficients of reflection and transmission be *r* and *t* respectively (for amplitudes); then conservation of energy requires that

$$r^2 + t^2 = 1 \qquad \text{...(10·14)}$$

The reflected wave has amplitude *ar* and transmitted wave amplitude *at*.

Now consider the reflected and transmitted waves as *reversed.* After reflection and refraction the total result must be a wave travelling along *OP* with amplitude *a*. Taking the coefficients of reflection and transmission for a wave incident from medium 2 as *r'* and *t'* respectively, we must have

$$att' + arr = a \qquad \text{...(10·15)}$$

and

$$art + atr' = 0 \qquad \text{...(10·16)}$$

These, together with Eq. (10·14) lead to

$$t' = t\ ;\ r' = -r \qquad \text{...(10·17)}$$

The negative sign in the second equation says that there is a *relative phase change* π in one case relative to the other. This equation does not tell us which of the two reflections involves the phase change π. For this we have to use the detailed work of Fresnel (§ 10·7); he showed that *reflection from the rarer medium side involves a phase change* π. This conclusion matches the string experiment.

10·8. Ripple tank Demonstration of Reflection and Refraction. An excellent demonstration of wave reflection and refraction is given with a ripple-tank using a straight dipper. One can show appearance of secondary wavelets, reflection at a hard boundary (showing $i = r$), and refraction at a plane edge or curved edge separating two 'media'. In the last case one part of the tank is shallower than the other, and since speed of waves depends on the depth (for small depths only), the two parts of the tank behave as different media.

We shall make no attempt to describe these experiments, because they are best tried and seen. One can verify Snell's law and also see a pulse gathering to form a real image or diverging out if a lens-shape is simulated.

PROBLEMS

10·1. Use Huygens' principle to explain total internal reflection.

10·2. The atomic structure of matter is neglected in this Chapter, and the medium is treated as continuous. Discuss the reasons why this is valid in case of visible light, but not in the case of x-rays.

10·3. Apply Huygens' principle to deduce the relation $1/v+1/u=2/r$ for image formation by a concave mirror. To what extent is this relation valid ?

10·4. Deduce the formula $1/v-1/u=(n-1)\ (1/r_1-1/r_2)$ for image formation by a concave lens, using Huygens' principle directly.

10·5. Using eq. (10·12) deduce the percentage reflection at water glass boundary for normal incidence from (*i*) water side, (*ii*) glass side, ($n_{glass}=1{\cdot}5$. $n_{water}=1{\cdot}3$).

10·6. Deduce the speed of light in a medium of $n=1{\cdot}60$. Also deduce the wavelength of mercury green line in this medium, given $\lambda_{vac}=5461$ $A°$.

10·7. At normal incidence an air-glass interface reflects about 4% intensity. If 10 glass plates are placed parallel, deduce the total loss of intensity in transmission through them. Now if the plates are put all in contact, keeping the gaps much smaller than λ, the loss would be just about 8%. Explain.

10·8. Name the scientists who stated that the refractive index n of a medium is equal to $\sin i/\sin r$, or c_{vac}/c_{med}, or c_{med}/c_{vac}. Which one of these is not supported by measurement and who made it first ? Is c_{med} used here the phase velocity or the group velocity ?

10·9. In § 10·3, instead of deducing the magnitude of AD, you could draw BD at an arbitrary angle r from AB, and then use that the *time* from any part of W_1W_2 to the corresponding part of BD is to remain *constant* if BD is the refracted wavefront. Use this method to prove the laws of refraction.

10·10. A simple proof of phase change π at one of the reflections follows from the fact that a film of thickness tending to zero should show zero reflection. Write down the formal proof.

11

Interference of Light (I)

When two or more waves arrive at a point in space simultaneously, the disturbance at that point is the *vector* sum of the disturbances caused by those waves taken individually. This is the general principle of *superposition of waves*, and applies to all kinds of waves. A special case of this arises when the superposing waves have the same frequency and cause disturbances along the *same line*. The result is then called *interference of waves*. In this and the next chapter we will study this phenomenon in relation to light. Two equal corpuscles (of Newton) reaching a point in space would always mean 2-fold intensity, but two equal waves (of Huygens) reaching a point in space can mean anything from 4-fold to zero intensity, depending on the phase relation between the two waves. That is the crux of interference.

We will first develop the general theory of interference and then apply it to the case of light.

11·1. Theory of Interference. Let the displacement y_1 at the observation point due to one wave at an instant t be given by

$$y_1=a_1 \cos 2\pi\nu t. \qquad \ldots(11{\cdot}1)$$

Let the second wave have the same frequency ν and cause displacement in the same direction. Then displacement y_2 due to it at the same point at the instant t may be represented by

$$y_2=a_2 \cos (2\pi\nu t+\delta), \qquad \ldots(11{\cdot}2)$$

where δ represents the *phase-difference* between the waves at the point under consideration.

Young's principle of superposition states that when both waves are present simultaneously the resultant displacement y at that point is given by*

$$y=y_1+y_2 \qquad \ldots(11{\cdot}3)$$

---

* Actually the principle of superposition is of wider application in which y_1 and y_2 could have different directions. Then eq. (11·3) is a vector equation. In discussing interference, we presently limit ourselves to cases where y_1 and y_2 are along the same line.

Substituting from eqs. (11·1) and (11·2) we work out

$$y=(a_1+a_2\cos\delta)\cos 2\pi\nu t-(a_2\sin\delta)\sin 2\pi\nu t$$

or $$y=a\cos(2\pi\nu t+\phi), \quad \text{...(11·4)}$$

where we can see that a and ϕ satisfy the relations

$$a^2=a_1{}^2+a_2{}^2+2a_1a_2\cos\delta \quad \text{...(11·5 a)}$$

and $$\tan\phi=\frac{a_1\sin\delta}{a_1+a_3\cos\delta} \quad \text{...(11·5 b)}$$

Let us examine eq. (11·5*a*). When phase difference δ is an even whole number of π, we get $a=a_1+a_2$. This is called the case of *additive interference.* On the other extreme, when δ is an odd multiple of π, we get $a=a_1-a_2$. This is called *destructive interference.* In fact, now a becomes 0 when $a_1=a_2$.

Example 11·1. Two waves of amplitudes 6 and 1 units are superposed with their vibrations parallel. Deduce the ratio of maximum to minimum intensity as the phase relation varies.

Solution. Maximum amplitude $=6+1=7$ units

Minimum amplitude $=6-1=5$ units

$$\frac{I_{max}}{I_{min}}=\frac{7^2}{5^2}=\text{nearly } 2.$$

Example 11·2. In the above case give a plot of intensity versus phase difference.

Solution. Using eq. (11·5*a*),

$$I=6^2+1^2+2\times6\times1\cos\delta=37+12\cos\delta$$

Fig. 11·1 shows the variation 12 cos δ superposed over a back-ground of 37 units.

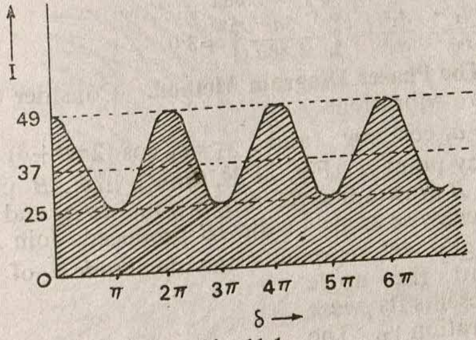


Fig. 11·1

For example 11·2

We could also put eq. (11·5*a*) as follows:

$$a^2=(a_1-a_2)^2+4a_1a_2\cos^2(\delta/2)$$

Thus we may treat the distribution as a superposition of variation $24\cos^2\delta/2$ over a background of 25 units. (Fig. 11·1).

If we have several waves reaching the observation point, we represent the disturbance caused by individual waves by

$$y_n = a_n \cos (2\pi\nu t + \delta_n) \qquad n = 1, 2, 3, \ldots$$

By the principle of superposition, we get

$$\begin{aligned} y = \Sigma y_n &= \Sigma\, (a_n \cos \delta_n \cos 2\pi\nu t - a_n \sin \delta_n \sin 2\pi\nu t) \\ &= C.\cos 2\pi\nu t - S.\sin 2\pi\nu t \\ &= A \cos (2\pi\nu t + \delta), \end{aligned} \qquad \ldots(11.6)$$

where $C = \Sigma a_n \cos \delta_n$ and $S = \Sigma a_n \sin \delta_n$, and further

$$A = \sqrt{C^2 + S^2}\,; \quad \tan \delta = S/C$$

Two extreme cases are of particular interest—

(*a*) When both $\Sigma a_n \cos \delta_n$ and $\Sigma a_n \sin \delta_n$ become zero, the resulant intensity is zero.

(*b*) When all the N waves reach in the same phase and have equal amplitude, then

$$A = Na_1\,; \qquad A^2 = N^2(a_1{}^2)$$

Thus the intensity is N^2 times the intensity due to an individual wave (not N times).

Example 11·3. Five vibrations of equal amplitudes are superposed first with all in phase agreement, and then with successive phase differences 40°. Calculate the ratio of resultant intensities in the two cases.

Solution. In the first case $\Sigma a_n \cos \delta_n = 5a$, and $\Sigma a_n \sin \delta_n = 0$

$$\therefore \qquad A_1 = [5a)^2 + 0^2]^{1/2} = 5a$$

In the second case

$$\Sigma a_n \cos \delta_n = a(1 + \cos 40° + \cos 80° + \cos 120° + \cos 160°) = 0.50a$$

$$\Sigma a_n \sin \delta_n = a\ (0 + \sin 40° + \sin 80° + \sin 120° + \sin 170°) = 2.84a$$

$$\therefore \qquad A_2 = a[0{\cdot}50^2 + (2.84)^2]^{1/2} = 2.88a$$

$$\therefore \qquad \frac{I_1}{I_2} = \frac{A_1{}^2}{A_2{}^2} = \left[\frac{5a}{2.88a}\right]^2 \approx 3{\cdot}0.$$

11·2. The Phaser Diagram Method. Consider the case of the addition of two vibrations

$$y_1 = a_1 \cos 2\pi\nu t \quad \text{and} \quad y_1 = a_2 \cos (2\pi\nu t + \delta)$$

From any point A (Fig. 11·2), draw a line AB proportional to a_1. Now from B draw a line BC proportional to a_2 and in a direction making anticlockwise angle δ with AB produced. Join AC.

The length AC represents the *amplitude a* of the resultant vibration, and the angle $CAB = \phi$ represents its *phase* relative to vibration y_1. The reader may verify that results are the same as those of eq. (11·5).

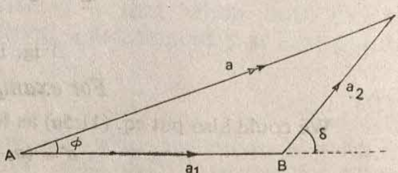


Fig. 11·2
Addition of two vibrations

Fig. 11·2 is called a *phaser diagram*, because the angles of BC and AC (relative to AB) represent phase differences, *not the directions* of vibration. In fact, the component waves cause vibrations along a common direction (say y-axis).

Fig. 11.3 gives the case of adding several simple harmonic vibrations of same frequency. We choose any reference phase line OX, and successively draw lengths OA_1, A_1A_2, $A_2A_3 \cdots$ proportional to the amplitudes of the constituent vibrations and in directions making angles with OX equal to their phase angles $\phi_1, \phi_2, \phi_3 \cdots$ The resultant vibration is shown *in amplitude and phase* by closing the last arm of the polygon taken in opposite sense. In this drawing all ϕ's must be measured anticlockwise from OX. Thus ϕ_4 (not shown) is greater than π.

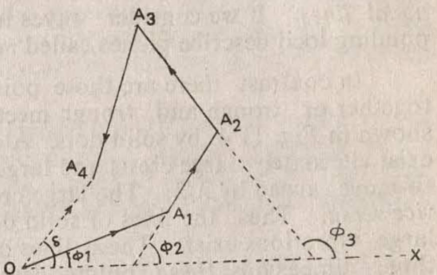


Fig. 11·3.
Addition of several vibrations

If the N vibrations thus added together close the polygon, the resultant is a vibration of zero amplitude.

11·3. Interference in Space : Nodal and Anti-nodal Planes. Consider two point sources S_1 and S_2 (fig 11.4) emitting waves of wavelength λ in the *same phase*. For discussion let them be two dippers on a water surface, generating crests and troughs. For simplicity we assume S_1 and S_2 as sources of equal strength.

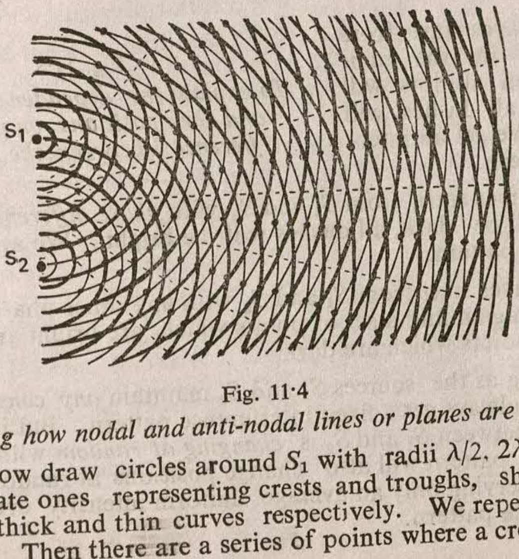


Fig. 11·4
Showing how nodal and anti-nodal lines or planes are formed

We now draw circles around S_1 with radii $\lambda/2, 2\lambda/2 \ldots n\lambda/2 \ldots$, the alternate ones representing crests and troughs, shown in the figure by thick and thin curves respectively. We repeat the same around S_2. Then there are a series of points where a crest from one source meets a trough from the other. The locii of these points are marked by the dashed lines. At these points there is neither upward nor downward displacement. If we let the waves moves $\lambda/2$ ahead, the crests become troughs and troughs become crests, but at these points the total displacement is still zero. Thus the dashed lines represent positions of zero vibration. These lines are called *nodal lines.* If we consider waves in three dimensions the corresponding locii describe planes called *nodal planes.*

In contrast there are those points where crest and crest meet together or trough and trough meet together. These points are shown in Fig. 11·4 by solid dots. Along the line joining them there exist alternately large crests and large troughs. Imagine the waves to move ahead by $\lambda/2$. The large crests turn to large troughs and *vice versa.* Thus the lines of solid dots represent positions where large vibrations exist. These lines are called *anti-nodal lines.* In three dimensions these points describe planes called *anti-nodal planes.*

If the sources S_1 and S_2 vibrate in opposite phase, the nodal and antinodal planes interchange.

11·4. Coherent Sources. Consider two sources S_1, S_2 (Fig. 11.5) emitting light of wavelength λ. On a screen AB, the dots represent points for which the distances from S_1 and S_2 differ by $n\lambda$, where n is any integer, and the dashes correspond to path differences $(n+\frac{1}{2})\ \lambda$. Now—

Let waves start from S_1 and S_2 in the *same phase.* Then all dots will be bright, all dashes will be dark. We get an interference pattern.

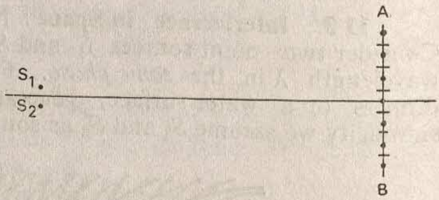


Fig. 11·5

To explain the dependence of bright and dark positions on phase difference δ between the sources

Let waves start from S_1 and S_2 with *phase difference* π. Then all dots will be dark, all dashes will be bright. We again get an interference pattern.

Let waves start from S_1 and S_2 with any other phase difference δ. We can again find a set of points which are bright and an alternate set of points which are dark.

So long as the sources S_1 and S_2 maintain *any constant phase-relation*, we always get a fixed interference pattern. But if the phase difference between S_1 and S_2 is *changing at random* with time, the interference pattern will also change positions at random and we shall be observing only an average uniform intensity everywhere—no inteference pattern.

Two sources, such as S_1 and S_2 *which maintain any constant phase relatian between themselves are called coherent sources.*[*]

In the case of mechanical vibrations or electronic oscillators we can control two independent sources and maintain a constant phase relation between them. But the emission of light in any lamp comes from billions of atoms, the emission from each one of them being random. *Hence two independent sources of light can never be coherent.* In other words, interference of light can never be observed with two independent sources.

If *both sources of light are obtained from a single parent source by some device,* then all the random phase changes occurring in the parent source are repeated in the secondary ones also, so that relative phase remains constant. These are called coherent sources of light.

11·5. Conditions for Interference of Light. We are now in a position to state the conditions for interference of light :

(1) *The sources must be coherent, i.e.*, they must keep a constant phase relation.

(2) *The beams must be propagated along nearly the same line* otherwise, their vibrations cannot be along a common line.

(3) *If the beams are polarized, their plane of polarization must be the same.* (This means that the electric vectors E in the interfering waves must be parallel.)

(4) *The sources must be of nearly equal intensities.*

(5) *The common original source must be monochromatic, i.e., emitting light of a single wavelength.*

Out of these conditions, (1) to (3) are fundamental, that is, interference cannot at all be observed without them. Conditions (4) and (5) are only matters of convenience in observations; they are not necessary.

Let us examine condition (4). If the two interfering sources have intensity ratio 100 : 1, their amplitude ratio is 10 : 1. Hence, the maximum to minimum amplitude ratio in the inteference pattern is (10+1) : (10−1), *i.e.*, 11 : 9, leading to intensity ratio $11^2 : 9^2$, *i.e.*, 121 : 81 or nearly 3 : 2, It is not very hard to judge this variation of intensity. Thus interference patterns *can* be seen *easily* even with two sources of intensity ratio 100 : 1.

Regarding condition (5), it is worth stating that interference patterns *are* formed for each of the wavelengths in a heterogenous source. Consider a mercury lamp source, which gives prominent spectral lines in red, yellow, green and violet. If we set, for instance, the biprism arrangement (§ 11·9) and then look for the fringes with a red, yellow, green or violet glass before our eye, then in each case we will find the fringes present.

---

[*] Note that S_1 and S_2 being mutually coherent does not necessarily mean that they are in the same phase. It only means that S_1 and S_2 have a constant phase difference.

Example 11·4 **In an interference pattern with two coherent sources the intensity variation is found to be 5% of the average intensity. Deduce the relative intensities of the interfering sources.**

Solution. Let the *amplitude* ratio of the sources be $a : 1$. $(a>1)$

Then in the interference pattern the intensity ratio is

$$(a+1)^2 : (a-1)^2$$

In the present case,

$$\therefore \quad \frac{(a+1)^2}{(a-1)^2}=\frac{105}{95}, \quad \Rightarrow \quad a\approx 40 \text{ (check)}$$

$\therefore$ *Intensity* ratio of the sources$=a^2 : 1=1600 : 1$ ⊙

11·6 Theory of Two Slit Interference. In Fig. 11·6 let S_1, S_2 represent two *coherent** slit sources the slit lengths being perpendicular to the plane of paper) giving out light of wavelength λ. Distance $S_1S_2=2d$. Consider a point P on a plane at distance $D >> 2d$. Let $OP=y<<D$. Then path difference $S_2P-S_1P=p$ varies with y and we get alternate maxima and minima on the plane OP

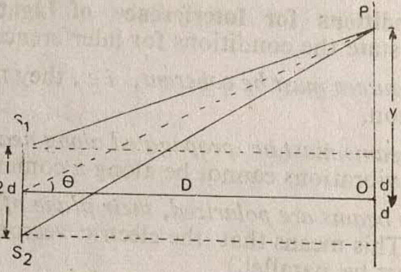


Fig. 11·6

For deducing fringe-width in interference

Since we have $d<<D$ and $y<\leqslant D$, we get

$$p=S_2P-S_1P\approx 2d.\,\theta\approx \frac{2yd}{D} \qquad \ldots(11.7)$$

Now, if S_1 and S_2 be in the same phase, then P will have *maximum* intensity if P is a whole number of wavelengths, *ie.*,

$$\text{for MAXIMA} \quad \frac{2yd}{D}=n\lambda, \qquad n \text{ an integer.} \qquad \ldots(11{\cdot}8)$$

Also P will have *minimum* intensity if p is a whole number plus half wavelength, *i.e.*,

$$\text{for MINIMA} \quad \frac{2yd}{D}=(n+\tfrac{1}{2})\lambda, \qquad n \text{ an integer.} \qquad \ldots(11{\cdot}9)$$

---

* We will describe later how coherent sources are obtained.

Thus as y changes, the conditions for maxima and minima are alternately satisfied. So we get alternate bright and bark *points* in the plane OP. If we consider the figure moved perpendicular to the plane of paper, we visualise alternate bright and dark *lines* on the screen. These are called the *interference fringes.* n in eq. (11·8) is called the *order of interference* at the particular fringe.

If S_1 is ahead of phase by $\triangle\phi$ relative to S_2, the left sides in Eqs. (11·8) and (11·9) become $(2y\, d/D+\triangle\phi\, (\lambda/2\pi)$.

If y_n and y_{n+1} describe the positions of two successive maxima, then, using eq. (11.8) or (11·9), their separation ω is given by

$$w=y_{n+1}-y_n=(n+1)\frac{\lambda D}{2d}-(n)\,\frac{\lambda D}{2d}=\frac{\lambda D}{2d} \qquad \text{...(11.10)}$$

The separation of any bright fringe from the nearest dark fringe comes out as just half of this. Thus we have alternate bright and dark fringes at equal separation $w/2$. (Even if the sources S_1S_2 have a phase difference $\triangle\phi$, we have the same w.) This quantity w is called the *fringe-width*, and we note that

(1) w increases in direct proportion to D.

(2) w increases in inverse porportion to d.

(3) w increases in direct proportion to λ.

Example 11.5. If $\lambda=5893\ A^\circ$ (Sodium D light), $2d=1.00$ mm, $D=50$ cm, calculate the fringe-width.

Solution. $w=\dfrac{5893\times10^{-8}\times50}{1\times10^{-1}}=0{\cdot}295$ mm. ⊙

Example 11·6. In a particular two-slit interference pattern with $\lambda=6000\ A^\circ$ the zero-order and 10th order maxima fall at micrometer readings 12·34 mm and 14·73 mm. If λ is changed to 5000 A° deduce the positions of the zero-order and 20th order fringes, other arrangemens remaining the same.

Solution.

Zero-order fringe is independent of λ. So it falls at 12·34 mm with $\lambda=6000\ A$ is

$$w=\frac{14{\cdot}73-12.34}{10}=0{\cdot}239 \text{ mm.}$$

For $\lambda=5000\ A^\circ$, we get the new fringe-width

$$w'=0{\cdot}239\times\frac{5000}{6000}=0{\cdot}199 \text{ mm.} \Rightarrow 20w'=3{\cdot}98\text{mm}$$

∴ Position of 20th order fringe with $\lambda=5000\ A^\circ$ is

$$12{\cdot}34\pm3{\cdot}98=16.32 \text{ mm, } 8.36 \text{ mm.}$$ ⊙

The phase difference δ for disturbance reaching P (fig. 11·6) from S_1 and S_2 is proportional to y. If $\delta=k.y$, eq. (11·5 a) gives the intensity* of the resultant disturbance as

$$I=I_1+I_2+2\sqrt{I_1I_2}\cos ky$$

This tells us that as y increases, the intensity rises and falls following a cosine law.**

Average intensity in the interference pattern may be evaluated by integrating Idy over one fringe width. It is left to the reader to show that $I_{av}=I_1+I_2$. Thus interference does not result in net increase or loss of energy ; it only causes *redistribution* of energy.

11·7. Young's Arrangement for Observing Interference. The earliest demonstration of interference of light was given by Thomas

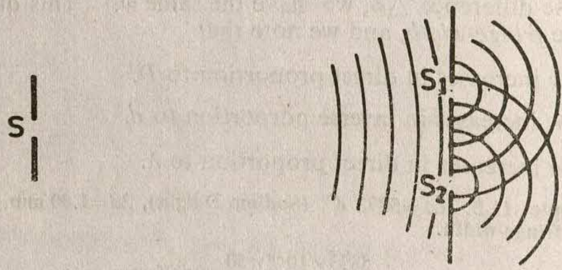


Fig. 11·7

Young's experiment

Young in 1801 with the arrangement shown schematically in fig. 11.7. S, S_1 and S_2 are three slits. The source is placed to the left of S. Both S_1 and S_2 receive waves from S and become the *coherent* secondary sources for waves to the right hand side of S_1 S_2. Interference fringes are observed there.

For success of the experiment S has to be narrow, and so also S_1 and S_2. In that case intensity is too weak. Hence this experiment is not easy to repeat and is now only of historical importance.

---

* Taking intensity proportional to (amplitude)^2.

** See also Example 11.2.

11·8. A Simple Equivalent to Young's Experiment. Press two safety razor blades together, and run their edges together on a blackened glass plate, like an exposed photographic plate. This produces two transparent scratches, the separation of which is nearly equal to the blade thickness. These scratches are the 'slits' S_1, S_2 shown in Fig. 11·8. Along a small length of these 'slits' open out a wider area, which we call the 'window' *W*.

Now, in the opaque cover of a sodium lamp let there be a long slit *S* about 1-2 mm wide (Fig. 11·9) and a number of small pin holes like *P* at distances about 2 cm apart. Standing about 2 to 5 metres away, look at *S* through the double-slit-and-window device held in front of the eye. Through the window *W* (Fig. 11·8) the slit *S* and holes *P* are seen as such, but through the pair of slits S_1, S_2 you see not one slit *S*, but a number of alternate bright and dark bands parallel to *S*. These are the interference fringes. If the eye position behind the device is so adjusted that the view is partly through S_1,S_2 and partly through *W*, then you see the fringes and also the bright holes.

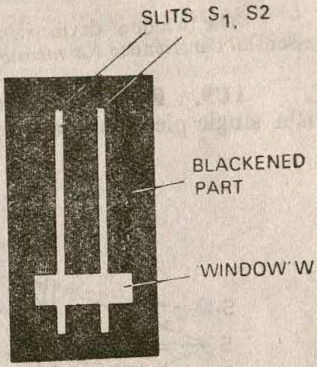


Fig. 11·8.

A magnified view of the 'slits' and window' scratched out on a blackened glass plate. Separation S_1S_2 *is* $\sim 0{\cdot}3$ *mm*

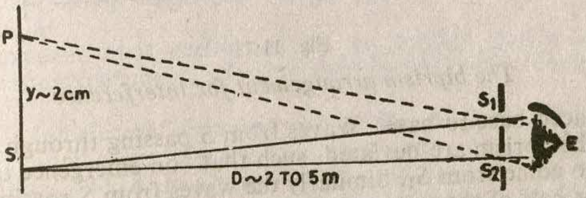


Fig. 11·9.

Schematic of the viewing arrangement. S_1S_2 *is a few mm, while D is few metres*

Fig 11·10 shows a typical view. The hole, seen at *B* in Fig. 11.10, may fall on any bright or dark fringe, or in-between, depending on the distance of observation *D* (Fig. 11·9), which is to be adjusted.

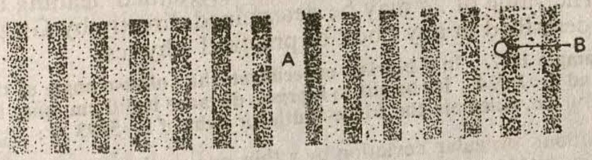


Fig. 11·10

A typical view in Young's experiment as described

If distance *SP* (fig. 11·9) is *y* cm and *n* fringes are seen between *A* and *B* (Fig. 11·10), the fringe width is *y*/*n*. This equals *w* of eq. (11·10), so that

$$w = \frac{y}{n} = \frac{\lambda D}{2d} \qquad \ldots(11{\cdot}11)$$

where $2d$ is the separation of the slits S_1S_2. We can measure blade thickness with a screw gauge and put $2d$ equal to this thickness. This will be true within ~20%. Since y, D and n are measurable, we get a close enough estimate of λ.

The 'window' device was first introduced by the author in 1969, and is an essential contrivance for *measurements.*

11·9. Fresnel's Biprism Arrangement. The *biprism* (Fig.11·11) is a single piece of glass, effectively equal to two prisms of narrow angle placed base-to base. Waves from S passing through the upper half of the prism are deviated such that on emergence the waves appear to come from S_1. Similarly the waves from S passing through the lower half of the prism deviate so as to appear coming from S_2. Thus we have *two coherent virtual sources* S_1, S_2. In the cross-shaded region in Fig. 11·11 the two coherent beams overlap and therefore interfere.

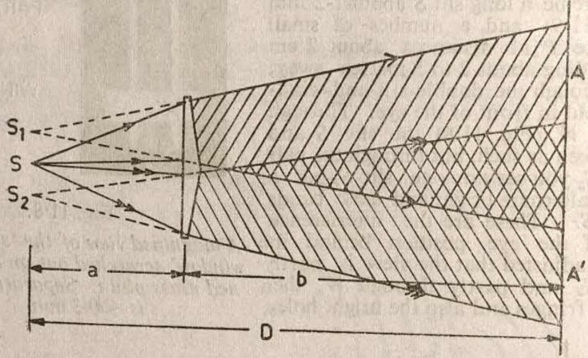


Fig. 11·11

The biprism arrangement for interference

An eyepiece placed in this region sees the interference fringes in its focal plane. If the eyepiece is capable of motion transverse to length D, the fringe-width w can be determined leading to λ (Eq. 11·10). The value of $2d$ may be determined directly by Glazebrook's method, described in any book of practical physics.

Example 11·7. In a biprism experiment, the obtuse angle of the biprism is 178°0′ and μ=1·524. If the slit to biprism distance is 5·0 cm and biprism to screen is 65·0 cm, calculate the fringe-width for light of λ=5893 $A°$.

Solution. Angular deviation by a thin angle prism is $(\mu-1)\,\theta$. For the biprism (Fig. 11·11) we get

$$2d = S_1S_2 = a\,(\mu-1)(\theta_1+\theta_2) = (5{\cdot}0)(0{\cdot}524)\left(2\times\frac{\pi}{180}\right) = 0{\cdot}091\text{ cm}$$

Here $\theta_1+\theta_2 = 180°-178° = 2° = 2\pi/180$ radian.

$$w = \frac{\lambda D}{2d} = \frac{5893\times10^{-8}\times(5+65)}{0{\cdot}091}\text{ cm} = 0{\cdot}045\text{cm}$$

⊙

11·10. Lloyd's Single Mirror Arrangement. In this device (Fig. 11·12) the original slit S and its image S' formed by a 'mirror' M act as the two coherent sources. The 'mirror' really is just a plain (not silvered) plate of glsas.* The region over which the coherent beams overlap is clearly shown in the figure by cross-shading. It can be readily seen that

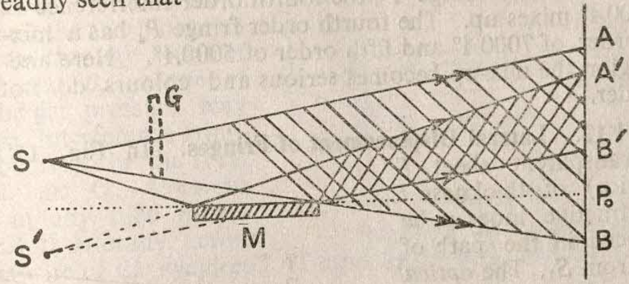


Fig. 11·12

The Lloyd mirror arrangement for interference.

$$2d \equiv SS' = a.\ 2\theta, \qquad \ldots(11{\cdot}12)$$

where θ is the grazing angle (90°−angle of incidence) at the centre of the mirror M.

The experiment is performed on an optical bench like the Fresnel biprism experiment, $2d$ being determined by the Glazebrook method.

11·11. Interference Fringes with White Light. White light, visible to the human eye, consists of a continuous range of wavelengths from about 4000$A°$ to about 7500$A°$. In any of the arrangements like biprism, Llyod's mirror, etc., let us replace monochromatic light by white light. We recall that the bright fringe positions are given by

$$y_n = \frac{D\lambda}{2d} n, \qquad \ldots(11{\cdot}13)$$

where n is an integer called the *order of interference*. The zero order fringe lies at $y=0$ for all the wavelengths. But for other orders the fringes for different wavelengths do not coincide.

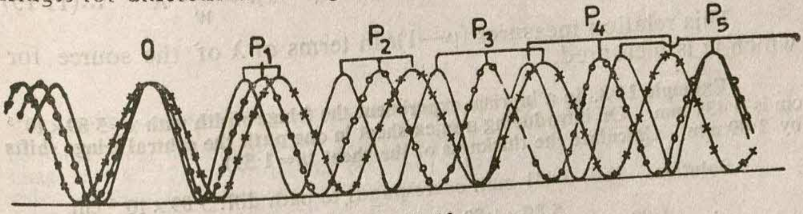


Fig. 11·13

Intensity distributions for three different wavelengths

Fig. 11·13 depicts the intensity distribution for three colours of λ ratio 5 : 6 : 7 (say 5000, 6000 and 7000A). The zero order maxima for all the three fall at P_0. This fringe is therefore white. In the region P_1, the nearer part has maximum for blue (5000*A*), the farther one for red (7000*A*°). This fringe has therefore a bluish tinge towards the central fringe and reddish tinge on the outer side. The next 'fringe' at P_2 shows these tinges more sharply. In the third order 'fringe' P_3 the fourth order interference maximum for 5000*A*° mixes up. The fourth order fringe P_4 has a mix-up from third order of 7000*A*° and fifth order of 5000*A*°. Here and beyond this region the mix-up becomes serious and colours do not follow any order.

---

* The plate is blackened at opposite face to avoid undesired light,

11·12. Lateral Displacement of Fringes. In Fig. 11·14, let a thin transparent sheet *T* (like mica) of thickness *t* and refractive index μ be introduced in the path of waves from S_1. The *optical* path for the wave from S_1 becomes longer by $(\mu-1)t$. If a point P_0' statisfies the relation

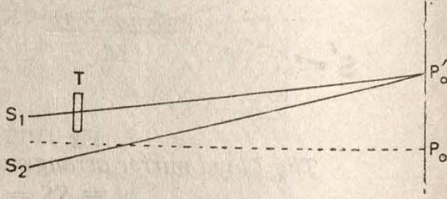


Fig. 11·14.

$$S_1P_0'+(\mu-1)t=S_2P_o'. \qquad ...(11\cdot14)$$

then the zero-order fringe (central fringe) will appear there. P_0P_0' $=y_s$ is called the *lateral shift* of the central fringe due to introduction of the thin sheet.

The zero order fringe (central fringe) can be identified with white light only; hence the shift P_0P_0' can be measured in white light fringes only.

Now suppose that *without altering any of the arrangements*, we just replace the white source with a monochromatic source of wavelength λ, and measure the fringe-width w. Then the lateral shift y_s measured above is equivalent to y_s/w fringes of wavelength λ, *i.e.*,

$$(S_2P_0'-S_1P_0')=(\mu-1)t=\frac{y_s}{w}\cdot\lambda \qquad ...(11\cdot15)$$

This relation measures $(\mu-1)t$ in terms of λ of the source for which w is measured.

Example 11·8. In a biprism experiment the fringe-width with $\lambda=5\cdot89\times10^{-5}$ cm is 0·431 mm. On introducing a mica sheet in one path the central fringe shifts by 1·89 mm. Calculate the thickness of the sheet ($n=1\cdot59$).

Solution. Shift 0·431 mm. corresponds to path diff. $5\cdot89\times10^{-5}$ cm

$\therefore$,, 1·89 ,, ,, $\dfrac{5.89\times1\cdot89}{0\cdot431}\times10^{-5}$ cm

Equating this with $(n-1)t$, we get

$$t=\frac{5\cdot89\times1\cdot89}{0\cdot59\times0\cdot431}\times10^{-5}\text{ cm.}=4\cdot4\times10^{-4}\text{ cm.}$$

[Note that the sheet introduces a path diff. $\approx 2{\cdot}6\times10^{-4}$ cm, and the lateral shift observed is $\approx 1{\cdot}9\times10^{-1}$ cm. In this sense the effect appears about 700 fold magnified !]

11·13. Rayleigh Refractometer. For measuring refractive index of a gas Rayleigh devised the arrangement whose principle is shown in Fig. 11·15. Slits S, S_1 and S_2 make Young's double-slit device, and the two coherent beams pass through separate tubes A and B, in one of which the gas pressure may be varied. Interference fringes are observed in the focal plane of lens L_2. Pressure change in any tube makes fringes shift laterally across the cross-wire of the eyepiece. If tube length is L, we have

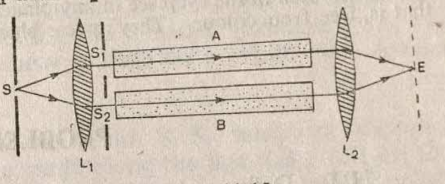


Fig. 11·15.

Rayleigh Refractometer

$$L\Delta\mu = m\lambda$$

where m is the number of fringes passing. The refractive index for any pressure can thus be obtained, since $\mu-1$ is proportional to the pressure. It may be noted that the tubes have to be about a metre long (unlike the very thin sheet T in Fig. 11·14) since $\mu-1$ for gas is $\sim10^{-4}$.

11·14. Achromatic White Light Fringes. If we arrange that the fringe-width is independent of λ and zero order fringe for all wavelengths is common then *all* fringes are free from colour. This fringe system is said to be achromatic. Lloyd mirror arrangement provides such a possibility. Let the 'source'

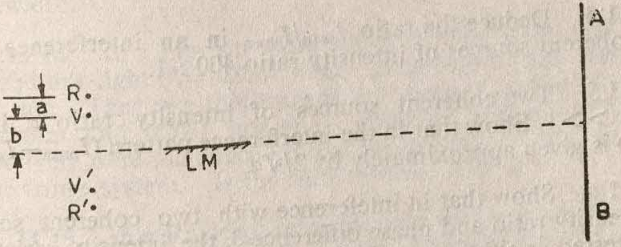


Fig. 11·16

Arrangement for achromatic fringes. LM is the Llyod mirror

be not a usual illuminated slit, but a short spectrum VR (Fig. 11·16). whose image is $V'R'$. We note that the perpendicular bisectors of V,V' and R,R' coincide, so that the zero order fringe on AB has the same position for V (violet) as for the R (red) end of the spectrum.

Let the width of spectrum RV be a, and let the distance of V from the plane of the mirror be b. Then it can be seen that

$$VV'=2b,\ RR'=2(a+b)$$

We can alter b by abjusting the tilt of the mirror and so arrange that

$$\frac{2d_{red}}{2d_{violet}}=\frac{RR'}{VV'}=\frac{2(a+b)}{2b}=\frac{\lambda_{red}}{\lambda_{violet}} \qquad ...(11\cdot16)$$

This makes ω(red)=ω (violet), *i.e.*, the fringe width is independent of colour. The fringes seen in the eyepiece in any plane like *AB* are therefore achromatic, that is, free from colour. They appear black-and-white.

PROBLEMS

11·1. Define *coherent* sources. Explain why waves from two independent tuning forks of same frequency show interference, but not waves from two independent sources of light emitting the same kind of light.

11·2. Two vibrations of equal amplitude *a* are superposed. Compute the amplitude *A* of the resultant for various values of phase difference δ between them, ranging from 0° to 360° in steps of 30°. [Phaser method may be very convenient]. Draw a graph showing (*i*) *A* against δ, (*ii*) A^2 against δ.

11·3. Deduce by the phaser method the resultant amplitude of vibration when six vibrations of equal amplitudes *a* are superposed with successive equal phase differences (*i*) 10°, (*ii*) 30°, (*iii*) 60°, (*iv*) 90°.

11·4. Deduce the ratio I_{min}/I_{max} in an interference pattern with coherent sources of intensity ratio 400 : 1.

11·5. Two coherent sources of intensity ratio α interfere, where $\alpha>>1$. Show that in the interference pattern $(I_{max}-I_{min})/I_{max}+I_{min})$ is given approximately by $2/\sqrt{\alpha}$.

11·6. Show that in inteference with two coherent sources of any intensity ratio and phase difference δ, the intensity variation can be expressed as either a cos δ variation or a $\cos^2(\delta/2)$ variation.

11·7. Show that if *n* sources of equal intensity send out waves at a point with phase 0, ϕ, 2ϕ...$(n-1)\,\phi$, the resultant oscillation is zero if $n\phi$ is an integral multiple of 2π.

11·8. Fig 11·4 is only a small part of the total drawing around S_1S_2. Draw a complete drawing extending to 10λ on all sides. Comment on the nodal lines ; also on the nodal planes if S_1 and S_2 are point sources sending waves in three dimensions.

11·9. Two coherent beams of wavelength 5000 *A*° reaching a point would individually produce intensities 1·44 and 4·00 units.

If they reach there together the intensity is 0·90 units. Deduce the lowest phase difference with which the beams reach that point.

11·10. Interference of electromagnetic waves is demonstrated with microwave oscillators S_1 and S_2. If $\lambda=4{\cdot}0$ cm, $S_1S_2=20{\cdot}0$ cm, and plane of observation is parallel to S_1S_2 and 200 cm away compute the separation of maxima and minima. Comment why two microwave oscillators may show interference while two sodium lamps (for example) do not.

11·11. Two microwave oscillators S_1,S_2 send out waves of $\lambda=8{\cdot}5$ cm. Observations are made along the line S_1S_2. Show that for fixed S_1S_2 distance there are no alternate maxima-minima along this line. Also explain what happens if distance S_1S_2 is varied from 10·0 cm to 30·0 cm.

11·12. A Fresnel biprism arrangement is set and in the field of view of the eyepiece we get 62 fringes with sodium light ($\lambda=5893$ $A°$). How many fringes shall we get in the same field of view if we replace the source with a mercury lamp using (*a*) green filter ($\lambda=$ 5461$A°$), (*b*) violet filter ($\lambda=4358A°$) ?

11·13. Briprism fringes are produced with the following specifications : Source : sodium yellow ($\lambda=5893$ $A°$), biprism : $\mu=1{\cdot}50$ and refracting angles 1·04° and 1·23°, distance from slit to biprism 12·4 cm and distance from slit to focal plane of the eyepiece 68·5 cm. Calculate (*a*) the separation of the coherent slit images, (*b*) the fringewidth.

11·14. In a biprism experiment fringes were first observed with sodium light ($\lambda=5893A°$) and fringe-width was measured as 0·347 mm. Then two thin transparent sheets *A* and *B* of thicknesses 0·016 mm and 0·020 mm and refractive indices 1·65 and 1·45 respectively were introduced in the two beams. Calculate the shift of the fringe system. Is the shift towards *A* or *B* ?

11·15. In a Rayleigh refractometer with cell length 80·0 cm a pressure change 10·0 cm of *Hg* causes 73 fringes of 5461 *A* to pass across the observing eyepiece. Compute μ of the gas at pressure 76 cm.

11·16. In a certain region of interference we get the 492th order *maximum* for sodium 5889. 97 $A°$ line. What will be the order of interference at the same place for (*i*) Sodium 5895·93 $A°$, (*ii*) Mercury 5460·7 $A°$, and (*iii*) Hydrogen 4861·4 *A* ?

11·17. A radio transmitter 10·0 m above an airfield emits radiations at $\lambda=80{\cdot}0$ cm. Interference between the transmitter waves and those from its reflection on the conducting ground produces one of the maxmium signals at angle θ degrees from the air field. Compute θ, as also the altitude of an aeroplane which receives such a signal and is 20 km away from the airfield. [This is the principle of 'glide path' in aircraft landing systems.]

11·18. Biprism fringes are observed with white light using a band-pass filter. How many fringes will be visible if the band range is (*i*) 5000 to 5500 *A*°, (*ii*) 4500 to 6000 *A*° ?

11·19. Explain the following

(*a*) Achromatic white light fringes can be formed with Lloyd mirror arrangement, but not with biprism arrangement.

(*b*) Using lateral displacement of fringes by introducing an extra path in one arm of a 2-stit interference system gives a very high effective magnification.

(*c*) If in the path from one of the coherent sources an extra length of ~20 cm in introduced, the interference fringes are usually not observed, though the sources are coherent.

12

Interference of Light (II)

12·1. Multiple Reflection : Division of Amplitude. In the cases of Young's experiment, Fresnel's biprism, Llyod's mirror, etc., the coherent sources are obtained by *dividing the wave-front* in width. Another way of producing coherent beams is to use *multiple reflection* between two faces of a transparent material. In this case widths of the successively transmitted as well as the reflected wave-fronts are about the same as that of the incident one ; but the amplitudes are different. A number of coherent beams—not just two—are produced on both sides of a film by a single incident beam, and these beams show interference. Thus, here we have coherent beams produced by *division of amplitude*. Note that the basis is that at any boundary between two different media there is always *part-reflection* and *part-transmission*.

A formula given by Fresnel (see § 10·5) shows that the *fraction of amplitude reflected* at the surface of separation is given by

$$r=\frac{n-1}{n+1} \text{ (for normal incidence)} \qquad \ldots(12{\cdot}1)$$

where n in the refractive index of the second medium relative to the first. For glass ($n=1{\cdot}5$), this gives $r=0{\cdot}2$. The fraction of amplitude transmitted (t) is given by the rule $t^2+r^2=1$, conserving energy. Hence we get for glass*

$$t=\sqrt{1-(0{\cdot}2)^2}=0{\cdot}98$$

Now consider a beam of amplitude a_0 (Fig. 12·1) successively reflected at the two faces of a film. The amplitudes of various beams shown are given by counting the number of reflections and transmissions suffered.

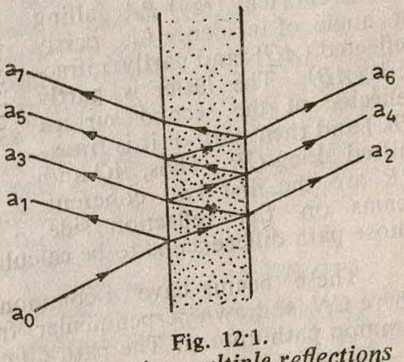


Fig. 12·1.
Showing multiple reflections

$$\begin{array}{l|l}
a_1=a_0r & a_2=a_0t^2 \\
a_3=a_0t^2r & a_4=a_0t^2r^2 \\
a_5=a_0t^2r^3 & a_6=a_0t^2r^4 \\
a_7=a_0t^2r^5 & a_8=a_0t^2r^6
\end{array}$$

---

* For water ($n=4/3$), we get $r=1/7$ and $t=0{\cdot}99$.

Thus, on the reflection side (left), the successive beams a_1, a_3, a_5 etc. have *amplitude ratios as follows* :

$$1 : t^2 : t^2r^2 : t^2r^4 \ldots\ldots \qquad \ldots(12{\cdot}2)$$

so that on the reflection side the *first two beams are of nearly equal amplitudes and the rest are of negligible amplitudes, since $t \approx 1$** and $r \ll 1$.

On the transmission side (right), the successive beams a_2, a_4, a_6 etc. have *amplitude ratios* as follows ;

$$1 : r^2 : r^4 : r^6 \ldots\ldots \qquad \ldots(12{\cdot}3)$$

so that the beams are of rapidly falling amplitude, since $r \ll 1$ *.

This contrast in the two cases explains the fact that interference *fringes in thin films are clear when seen in reflection, and very indistinct on the transmission side.**

Example 12·1. Partially reflecting films can be coated on the surfaces of transparent media. In one case the coating reflects 80% intensity. Neglecting absorption, compute the relative amplitudes of successive beams in reflection and in transmission.

Solution. Given $r^2 = 0{\cdot}80$, hence $t^2 = 1 - 0{\cdot}80 = 0{\cdot}20$
Using Eq. (12·2) the successive reflected beams have amplitude ratios

$$1 : 0{\cdot}20 : 0{\cdot}16 : 0{\cdot}13 \ldots\ldots$$

The transmitted beams have amplitude ratios

$$1 : 0{\cdot}80 : 0{\cdot}64 : 0{\cdot}51 \ldots\ldots$$

12·2. Interference in Thin Films. (*a*) *Path difference between coherent beams.* If the thickness of a film is not rapidly varying, any small part of it may be treated as having parallel faces. Fig 12·2 shows a part of the film, where its thickness is *e*. A wavenormal (ray) *PA* falling at angle of incidence *i* is partly reflected (*AQ*) and partly refracted (*AB*). The latter is partly reflected at the second surface (*BC*) and then part of it is transmitted along *CR*. Thus, *AQ* and *CR* are the first two coherent beams on the reflection side whose path difference is to be calculated.

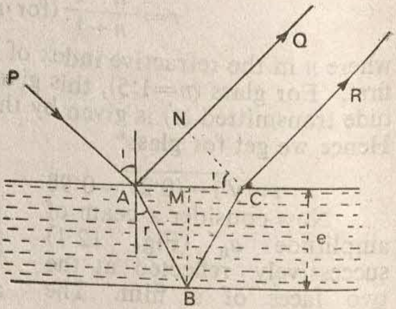


Fig. 12·2
Path difference in successively reflected beams

These beams have a common path beyond the plane *CN*, where *CN* is drawn perpendicular from *C* on *AQ* ; also they had a common path upto *A*. The path difference is, therefore, given by

---

* The situation changes if the films are *coated* to increase *r* (hence decrease *t*) See Example 12·1.

$$p=(AB+BC) \text{ in the film}-AN \text{ in air} \qquad \ldots(12\cdot4)$$

$$=n(AB+BC)-AN,$$

where n is refractive index of the film. Now, by simple geometry :

$$AB=BC=\frac{e}{\cos r}$$

$$AN=AC \sin i=2AM \sin i=2e\,\frac{\sin r}{\cos r}\cdot n \sin r$$

Substituting in eq. (12·4),

$$p=\frac{2n\,e}{\cos r}-\frac{2ne \sin^2 r}{\cos r}=\frac{2ne\,(1-\sin^2 r)}{\cos r} \qquad \ldots(12\cdot5)$$

$$=2ne \cos r.$$

Note that in eq. (12·5), r is the angle of refraction.

For normal incidence $p=2ne$. In fact for r as large as 20° ($\approx 1/3$ radian) $\cos r$ differs from unity by$\sim5\%$ only.

(*b*) *Fringes as Locii of Equal Thickness.* Consider the viewing of different parts of a film by an observer. The change in p is given by

$$\triangle p=2ne\triangle\,(\cos r)+2n \cos r\,(\triangle e)\approx 2n \cos r\,(\triangle e) \qquad \ldots(12\cdot5\text{ a})$$

Usually, the area of the film seen at a time is small compared to the distance from which the film is seen. Hence variation of r is small, and in the case of thin films ($2ne\ll\lambda$), the contribution of the first term is negligible. The main variation in p is contributed by the variation of thickness e of the film.

Now, let P (fig. 12·3) be a point on the film for which thickness e_m is such that path difference between successively reflected waves is $m\lambda$: $\qquad \ldots(12\cdot6)$

$$p\equiv 2n\; e_m \cos r=m\lambda.$$

In part-reflection at the two surfaces of the film there is a relative phase change π, as shown by Stokes (see § 10·7). For this reason $p=m\lambda$ corresponds to phase difference $2m\pi$ $+\pi$ *in the case of reflected beams.* Hence, P will appear as a *dark* point. Now consider all those points of the film at which thickness is the same as that at P. Since r variation is not contributing noticeably, all these points will satisfy eq. (12·6) and hence will appear cark.* We thus get the condition for a dark *fringe* as the locus of points where film thickness satisfies eq. (12·6). With different values of m, we get different e_m, which define successive dark fringes, Q, R etc. in Fig. 12·3.

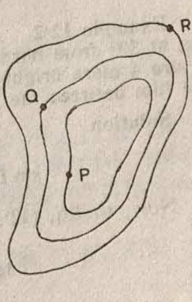


Fig. 12·3

Interference fringes as locii of equal thickness

---

* See, however, §12·4 (g).

Notice that *each fringe describes a locus of equal thickness of the film.* Obviously, no two fringes can ever cross. If we consider two points on successive fringes (Fig. 12·3), the difference of thicknesses of the film at these points is

$$\delta e = e_m - e_{m\pm1} = \frac{\lambda}{2n\cos r}\left(m - \overline{m\pm1}\right) = \pm\frac{\lambda}{2n\cos r}$$

If there are q fringes between any two points in the film and the viewing is near the normal,

$$\delta e = \pm\frac{q\lambda}{2n} \qquad \text{...(12·7)}$$

Thus just the counting of fringes gives the thickness difference.*

(*c*) *Fringes in a Wedge Film.* In a wedge film the locii of equal thickness are parallel to the edge of the wedge. Hence the fringes are parallel and equidistant straight lines. Fig. 12 4 shows a section of the wedge perpendicular to its edge, which is at E.

If the angle of the wedge is θ radian, and the fringes have a successive separation w (fringe width), we have

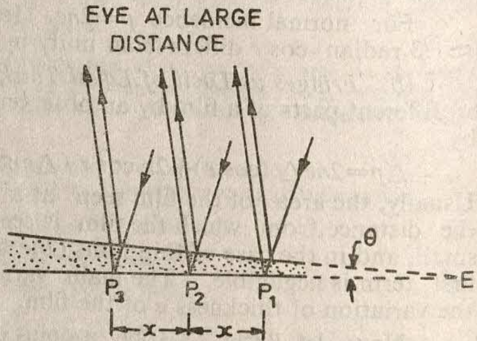


Fig. 12·4
A wedge film

$$|\,\delta e\,| = w\theta, \quad \Rightarrow \quad w = \frac{\lambda}{2n\,\theta} \qquad \text{—(12·8)}$$

Example 12·2. A soap film is seen in sodium yellow light with rays falling at 30° from normal. Bright fringes occur at two points P and Q and there are 4 more bright frınges in between. Deduce the difference of thickness of the film between the points. (n of soap film$=$1·33).

Solution.

$$\sin i = 0\cdot500, \;\therefore\; \sin r = \frac{\cdot500}{0\cdot33} = \cdot375 \Rightarrow \cos r = \cdot927$$

Now, in Eq. (12·5 a), $\delta p = 5\lambda$.

$$\therefore \quad \delta e = \frac{\delta p}{2n\cos r} = \frac{5\times5\cdot89\times10^{-5}\text{cm}}{2\times1\cdot33\times\cdot927}$$
$$= \mathbf{1\cdot19\times10^{-4}\ cm.}$$

Example 12·3. A wedge-shaped air-film between one fixed surface and another variable surface is seen with mercury green light ($\lambda = 5\cdot46\times10^{-5}$ cm).

---

* The fringes cannot decide between the + and— sign, unless some other information exists.

An eyepiece is fixed on a given fringe, and as the variable surface moves perpendicular to itself we find 47 fringes passing across the cross-wire. Deduce the change in position of the surface.

Solution. If initial fringe was of m^{th} order, final fringe at the same place has $(m\pm47)$th order. Assuming that the fringes are seen nearly normally $(r=0)$ and since the medium is air $(n=1)$, Eq. (12·7) gives

$$2\Delta e=\pm47\lambda=\pm47\times5{\cdot}46\times10^{-5}\text{ cm}$$

$$\therefore\quad \boldsymbol{\Delta e=\pm1{\cdot}3\times10^{-3}}\textbf{ cm.}$$

(Note that this represent a typical use of interference fringes in measuring small displacements due to elastic bending, thermal expansion, etc.) ⊙

Example 12.4. Between two optically plane glass plates, ***at one edge*** **a foil is introduced to get a wedge-shapee air film. Viewed in mercury green light ($\lambda=5{\cdot}46\times11^{-5}$ cm.) along the normal we see 12 fringes in 0·40 cm width. Deduce (*i*) the angle of the wedge; (*ii*) thickness of the foil if plate length is 3·0 cm in all, (*iii*) fringe width if water is introduced in the wedge space.**

Solution. Fringe-width $w=(0{\cdot}40/12)$ cm. Hence from eq. (12·8)

$$\theta=\frac{5{\cdot}46\times10^{-5}}{2\times1\times40/12}=\mathbf{8{\cdot}2\times10^{-4}}\textbf{ rad.}$$

Now θ=foil thickness $t\div$plate length L. Hence

$$t=8{\cdot}2\times10^{-4}\times3{\cdot}0=\mathbf{2{\cdot}5\times10^{-3}}\textbf{ cm.}$$

From eq. (12·8), w is proportional to $1/n$. Hence with water,

$$w'=\frac{0{\cdot}40}{12}\times\frac{1}{1{\cdot}33}=\mathbf{0{\cdot}025}\textbf{ cm.}$$ ⊙

(*d*) *Non-reflecting coatings*. An interesting use of interference of light is to make thin film coatings of a transparent material (like magnesium fluoride) on lens surfaces to reduce reflection losses (see § 10·5 and Example 10·3). The coating material is so chosen that $n_{air}<n_{coating}<n_{glass}$, so that the phase difference between the first two reflected beams is $2ne$ only (the π phase change being common). The thickness e is so chosen that*

$$2n_{coating}.e=\lambda/2$$

so that the first two reflected beams cancel out by interference. λ in the middle part of the spectral region of interest is chosen for best results. These are called *non-reflective coatings*.

12.3. Colours of Thin Films. When a soap film or an oil film on water is seen in *reflected* daylight it shows beautiful colours. The origin of these colours is explained satisfactorily on interference of multiply reflected waves of light.

Let the film have variable thickness, and let P, Q, R be three points on it with thicknesses satisfying :

$$\begin{aligned}2ne_1\cos r&=m\lambda_{red}\\2ne_2\cos r&=m\lambda_{green}\\2ne_3\cos r&=m\lambda_{blue}\end{aligned}\qquad\ldots(12{\cdot}9)$$

---

* With $\lambda\approx6000$ *A*, $n\approx1{\cdot}5$, and atomic diameters $\sim2A$. the films have to be ~400 atomic layers thick.

Here e_1, e_2, e_3 are the film thicknesses at P, Q, R ; and we consider inclination r as constant, since the region in view is small. Then at P the waves of red light interfere destructively (recall Stokes' phase change π)* ; the colour at P and *at all points of the same thickness* will therefore be bluish (white minus red). Similarly, the colour at R and at all points of the same thickness will be reddish (white minus blue) ; the colour at Q and all points of same thickness will be another tinge purple (white minus green), and so on.

These give the coloured fringes on the film. *Each fringe belongs to locus of a particular thickness of the film.* No two fringes can ever cross. Successive fringes of the same 'tinge' or colour represent change of thickness $\lambda/(2n \cos r)$, where λ is wavelength for the colour complimentary to that tinge. Further, from the order in which the colours show up we can tell which way the thickness is increasing.

Example 12·5. A uniform soap film has thickness $3{\cdot}0\times10^{-5}$ cm. What colour does it show when seen in reflected white light along the normal ?

Solution. For soap film $n=1{\cdot}33$. The wavelengths for *destructive* interference in reflected light are given by

$$2\times1{\cdot}33\times3{\cdot}0\times10^{-5}=m\lambda \qquad (m=1, 2\ldots)$$

$$\lambda=8\times10^{-5} \text{ cm}, 4\times10^{-5} \text{ cm}, 2{\cdot}7\times10^{-5} \text{ cm},\ldots\ldots$$

The possible λ values in the visible range are 8×10^{-5} cm (red) and 4×10^{-5} cm (violet). Hence these parts of the spectral colours are *absent* in the reflected light. Intermediate wavelengths correspond to yellow-green, and that colour will be seen.

Cases of a Thick Film and Infinitely Thin Film. If the film has large thickness, then, at a given point, the condition $2ne \cos r=m\lambda$ will be satisfied for several colours (with different m, of course), and no particular tinge will appear there. Consider a soap film ($n=1{\cdot}33$) of $e={\cdot}0005$ cm in near normal view. The path difference is equal to 20λ for red (6650 A, say) and 26λ for blue (5000 A, say). The condition $p=m\lambda$ is thus satisfied for 7 different colours from red to blue. This film hardly shows strong tinges of colour. Conversely, it follows that a film showing strong colours must be only a very few wavelengths thick.

On the other end imagine a film with thickness tending to zero. Then in reflected light the geometrical path difference between the coherent beams tends to zero. Taking Stokes' phase change π into account, the beams reflected by the two faces interfere destructively. The film therefore reflects no light ; all light is transmitted.

In fact we can reverse the argument. A film of thickness tending to zero leads to a situation of 'no film at all'. We cannot expect any reflection in such a situation. Hence a phase-change π between the two reflections is a theoretical necessity (see § 10·7).

---

* In case of an oil film on water, n (oil)$>n$ (water) for most oils. Hence for this case also phase change π occurs at one reflection and not at the other one.

12·4. Newton's Rings. These fringes, first observed by Newton, form only a particular case of fringes in thin films. But because they give an instructive laboratory experiment we shall discuss them in some detail.

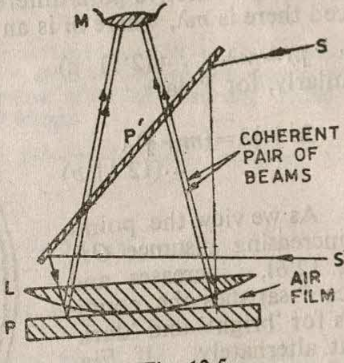


Fig. 12·5

The optical arrangement for Newton's rings

(*a*) *The essential optical arrangement.* It is shown in Fig. 12·5. *L* is a plano-convex lens of large focal length placed on a plane plate *P*, giving an *air-film of varying thickness* between them. Monochromatic light from a *wide source S* falls on this air-film by reflection from a 45° inclined plain glass plate *P'*. A microscope *M* focussed on the air-film shows a series of concentric bright and dark rings.

(*b*) *The coherent beams.* Fig. 12·6 shows in a magnified form how two coherent beams 2 and 3 are produced from a single incident beam 1 by reflections at the two faces of the air-film. If *e* is the thickness of air-film at distance *x* from the point of contact of the lens and plate, the path difference between the interfering beams there is 2*e*, and the effective difference is given by

$$p=2e+\frac{\lambda}{2} \qquad \ldots(12{\cdot}10)$$

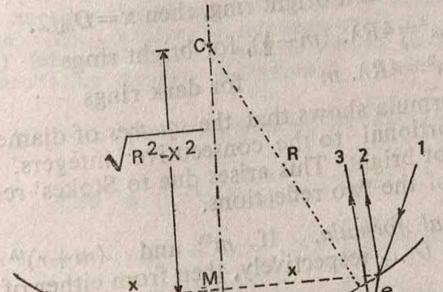


Fig. 12·6

The geometry in Newton's rings experiment

where $\lambda/2$ is added because of extra phase change π at the lower reflection (Stokes' result, § 10·7).

(*c*) *Bright and dark rings.* In Fig. 12·6, point P will appear bright if the effective path difference between coherent beams produced there is $m\lambda$, where m is an integer. Thus, for P bright :

$$2e+\tfrac{1}{2}\lambda=m\lambda \quad \ldots(12\cdot11\ a)$$

Similarly, for P dark :

$$2e+\tfrac{1}{2}\lambda=(m+\tfrac{1}{2})\lambda \quad \ldots(12\cdot11\ b)$$

As we view the points at increasing distance OP, (Fig. 12·6), e increases and hence P satisfies the conditions for 'bright' and 'dark' point alternately. If Fig. 12·6 is now imagined rotated about axis OC, we visualise a series of bright and dark interference rings with O as centre. Fig. 12·7 gives a schematic view of the fringes.

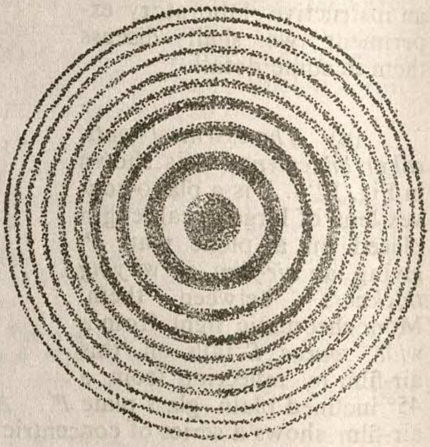


Fig. 12·7

Newton's Rings (Schematic)

(*d*) *Formula for diameters of Newton's Rings.* If R is the radius of curvature of the lower face of the lens, Fig. 12·6 gives

$$e=CO-CM=R-(R^2-x^2)^{1/2}=x^2/2R$$

using binomial theorem, since $x \ll R$. Substituting this in eq. (12·11 *a*), we get

$$x^2/R=(m-\tfrac{1}{2})\lambda \quad \text{for } P \text{ bright}$$

If D_m is the diameter of a bright ring, then $x=D_m/2$. Whence

$$D_m{}^2=4R\lambda.\ (m-\tfrac{1}{2}), \text{ for bright rings} \quad —(12\cdot12\ a)$$

Similarly, $$D_m{}^2=4R\lambda.\ m \quad \text{for dark rings} \quad —(12\cdot12\ b)$$

The last formula shows that the *squares* of diameters of dark rings are proportional to the consecutive integers. The centre ($D=0$) is dark, not bright. This arises due to Stokes' result of phase change π at one of the two reflections.

(*e*) *Practical formula.* If m^{th} and $(m+r)^{th}$ rings have diameters D_m and D_{m+r} respectively, then from either of eqs. (12·12)

$$D_{m+r}{}^2-D_m{}^2=4R\lambda.\ r$$

$$\therefore \quad \lambda=\frac{D_{m+r}{}^2-D_m{}^2}{4Rr} \quad \ldots(12\cdot13)$$

The formula remains unchanged whether we consider the bright or dark rings. The advantage of eq. (12·13) is that it applies even when the lens and plate do not touch at the centre (due to, say, dust particles).

Example 12·6. An equiconvex lens of focal length 80 cm (*n* of material 1·55) is placed over a plane plate. Newton's rings are formed with mercury green light ($\lambda=5{\cdot}46\times10^{-5}$ cm). Deduce (*i*) diameter of the first dark ring round the centre, (*ii*) difference of (Diameter)² for 10 fringes.

Solution. For an equiconvex lens, $1/f=(n-1).\ 2/R$. That gives $R=88$ cm.

Using eq. (12.12b) with $m=1$,

$$D_1=[4\times88\times5{\cdot}46\times10^{-5}\times1]^{1/2}=\mathbf{0{\cdot}14\ cm}\ \text{(check it)}$$

Using eq. (12·13) with $r=10$,

$$D_{m+10}^2-D_m^2=4\times88\times10\times5{\cdot}46\times10^{-5}=\mathbf{0{\cdot}19\ cm^2}\ \text{(check it).}$$ ⊙

(*f*) *Newton's Rings with a Liquid Film.* If a drop of liquid is introduced between the lens and plate of Newton's Rings device, it spreads out to give a liquid film instead of an air-film. The simplest way of taking this into account is to say that the wavelength of light changes from λ to λ/n, so that eq. (12·13) becomes

$$\frac{\lambda}{n}=\frac{(D_{m+r}^{\ 2}-D_m)^2{}_{\text{LIQ}}}{4Rr} \qquad \text{...(12·14)}$$

From Eqs. (12·13) and (12·14) we get

$$n=\frac{(D_{m+r}^{\ 2}-D_m^{\ 2})_{\text{AIR}}}{(D_{m+r}^{\ 2}-Dm^2)_{\text{LIQ}}} \qquad \text{...(12·15)}$$

This provides a method of determining *n* of a liquid. The method has the advantage of needing only a drop or two of the liquid.*

Example 12·7. In a certain set-up the Newton's rings observed with light of $\lambda=5{\cdot}89\times10^{-5}$ cm show the difference of squares of diameters of successive rings as 0·124 cm². What happens to this quantity if we (*i*) change λ to $4{\cdot}36\times10^{-5}$ cm. (*ii*) introduce a liquid of $n=1{\cdot}56$ between the lens and the plate, (*ii*) the lens surface facing the plate is changed to one of twice the radius o curvature.

Solution. All the results follow from eq. (12·14). The given quantity is

$$A\equiv D_{m+1}^{\ 2}-D_m^{\ 2}=\frac{4R\lambda}{n}=0{\cdot}124\ \text{cm}^2$$

For case (*i*) $A'/A=\lambda'/\lambda \Rightarrow A'={\cdot}124\times\dfrac{4{\cdot}36}{5{\cdot}89}=0{\cdot}092\ \text{cm}^2.$

For case (*ii*)$A'/A=n'/n \Rightarrow A'={\cdot}124\times\dfrac{1}{1{\cdot}56}=0{\cdot}079\ \text{cm}^2.$

For case (*iii*) $A'/A=R'/R \Rightarrow A'={\cdot}124\times\dfrac{2}{1}=0{\cdot}248\ \text{cm}^2.$ ⊙

---

* In practice, the intensity becomes very low when the liquid is introduccd. Consider eq. (12·1) for the case of water-and-glass interface. *n* is roughly $(3/2)\div(4/3)=9/8$. This gives $r=1/17$, compared with $r=1/5$ for air-film. Intensity falls by a factor $(5/17)^2\approx1/10$.

Example 12·8. Newton's rings are observed between a convex lens and a plane plate. The diameters of xth and x+5th rings are 11·37 and 14·28 units. Deduce the diameters of (*i*) x−5th, (*ii*) x+14th rings.

Solution. $\triangle (D^2)$ is proportional to the difference of the 'order' number. Here we have :

$$\triangle (D^2) \text{ for 5 rings} = 14{\cdot}28^2 - 11{\cdot}37^2 = 203{\cdot}9 - 129{\cdot}3 = 74{\cdot}6 \text{ (units)}^2$$

$$\therefore \quad D_{x-5}^2 = 129{\cdot}3 - 74{\cdot}6 = 54{\cdot}7\ ;\ \Rightarrow D_{x-5} = \mathbf{7{\cdot}40 \text{ units}}$$

$$\triangle (D^2) \text{ for 9 rings} = 74{\cdot}6 \times 9/5 = 134{\cdot}3 \text{ (units)}^2$$

$$\therefore \quad D_{x+14}^2 = 203{\cdot}9 + 134{\cdot}3 = 338{\cdot}2 \Rightarrow D_{x\times 14} = \mathbf{18{\cdot}40 \text{ units.}}$$

⊙

(*g*) *Newton's rings with a bright centre.* At the centre of Newton's rings the *geometrical* path difference is **0**, but due to Stokes' phase change π at the lower reflection (air to glass incidence), the centre is ordinarily dark. However, if the lens and plate are of different materials of refractive indices n_1 and n_2 and a liquid having refractive index *n lying between* n_1 and n_2 (see Fig. 12·8) is introduced between them, then both the reflections would be similar and the centre of Newton's rings should be bright. This was shown by Young, using oil of sassafras between crown glass and flint glass ($n_{flint} > n_{oil} > n_{crown}$).

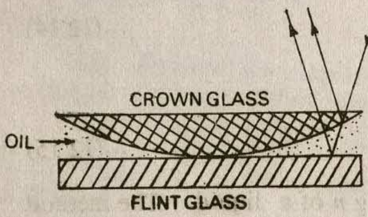


Fig. 12·8
Newton's rings with bright centre

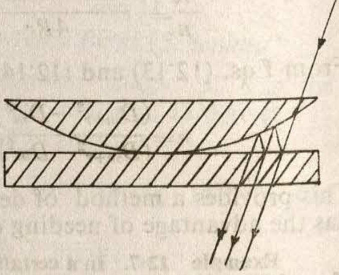


Fig. 12·9
Interference in transmitted light

(*h*) *The Fringes in Trasmitted Light.* On the transmission side successive beams differ from one another in amplitude ratio r^2 arising from *two* reflections within the film (See Eq. 12·3). Since r^2 is small, the contrast of maxima and minima is not good. For air-film in glass, considering two beams only, the intensity ratio from maxima to minima is $(1+.04)^2 : (1-.04)^2$, *i.e.*, a *variation* of about 16% only.

In Fig. 12·10 the curve on left shows a plot of intensity *I* in transmitted beam against D^2. The curve on right shows the variation of intensity in the fringes on the reflection side. The visibility of fringes depends *not* on intensity but on the contrast of I_{max} and I_{min} as given by

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \qquad \ldots(12{\cdot}16)$$

The visbility V is thus much larger for the reflected side than for the transmitted side.

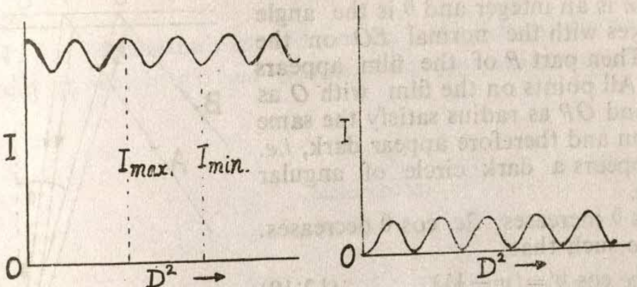


Fig. 12·10.

Intensity distribution (left) in transmitted light, (right) in reflected light

The Stokes' phase in *two* reflection (Fig. 12·9) is 2π, so that eqs. (12·12a) for bright and dark fringes interchange. Thus the intensity distributions on transmitted and reflected sides are complimentary.

12·5. Haidinger's Fringes in Parallel Films. The path difference $p=2\,ne\cos r$ between the coherent beams in multiple reflection from a film changes according to

$$\triangle p=2n\cos r.\triangle e+2ne.\triangle(\cos r) \qquad \ldots(12{\cdot}17)$$

At *small e* ($ne\approx$ *a few* λ) the second contribution is negligible compared with the first. Each fringe then represents the locus equal thickness. These are called *fringes of equal thickness* or Newton type fringes. So far we have been dealing with such cases only.

At *large e* ($ne\approx$ *a few thousands of* λ) the second contribution in $\triangle p$ becomes dominant; in a parallel film ($\triangle e=0$) that is the only contribution to $\triangle p$. Then each fringe corresponds to a given r. These are *fringes of equal inclination* or Haidinger's fringes.

To observe Haidinger's fringes the film thickness e must be appreciably large. For example, as r changes from 0° to 5° we have $\triangle(\cos r)=0{\cdot}004$; to get 40 fringes in this angular range with $n=1$ we need $2e\times\triangle(\cos r)=40\lambda$ or $e=5000\lambda$. This comes to $\sim$3 mm for yellow light. (In Newton's fringes, one gets 40 fringes for $\triangle e=20\lambda\approx12\times10^{-3}$mm).

The viewing of the fringes has to be different in the two cases. In Newton-type case the coherent beams from a given part of the film (but at different r values) are in similar state of interference. Hence the observing instrument is to be focussed *on the film*; it is to be a microscope. In Haidinger-type case the coherent beams for different parts of the film but ot the same r are to be collected by the observing instrument at one place; it is to be a telescope, focussed *at infinity*.

Fig. 12·11 shows the viewing of Haidinger's fringes with the naked eye E. Let P be such that

$$2ne\cos\theta = m\lambda \qquad \ldots(12{\cdot}18)$$

where m is an integer and θ is the angle EP makes with the normal EO on the film. Then part P of the film appears dark. All points on the film with O as centre and OP as radius satisfy the same condition and therefore appear dark, *i.e.* there appears a dark circle of angular radius θ.

As θ increases, $2e\cos\theta$ decreases. Let Q be such that

$$2ne\cos\theta' = (m-\tfrac{1}{2})\lambda \qquad \ldots(12{\cdot}19)$$

Then Q is bright and all points on a circle with O as centre and θ' as angular radius appear bright. Thus we get alternate bright and dark fringes.

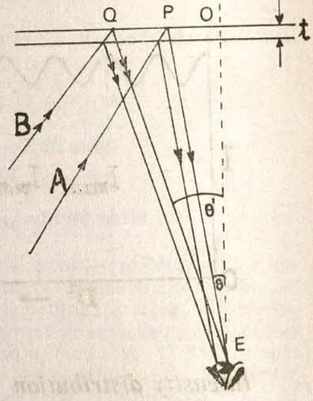


Fig. 12·11 *Observing Haidinger's fringes*

Note that position of the *centre* of the rings on the film depends upon the position of the eye ; the centre is at the foot of normal from the eye on the film. Note also that the linear size of the ring on the film depends on distance EO. The *angular size* alone is defined by the equations.

Viewed from the transmission side the conditions for dark and bright rings interchange (now no Stokes' phase change π), *i. e.*, the fringes are complimentary to those seen in reflected light.

Since θ is small, we may write for the m^{th} order fringe

$$2e\cos\theta_m = 2e\left(1-\frac{\theta_m^2}{2}\right) = m\lambda$$

whence

$$\theta_{m-1}^2 - \theta_m^2 = \frac{\lambda}{e}. \qquad \ldots(12{\cdot}20)$$

Thus differences of squares of *angular radii* of successive fringes are equal.* They decrease as e increases.

Example 12·9. A uniform plate of thickness 3·00 mm and $n=1{\cdot}50$ is viewed in light of $\lambda=6{\cdot}00\times10^{-5}$ cm. Calculate the (*i*) order of inteference m_0 at the centre, and (*ii*) angular radius of fringe of order m_0-10.

Solution. $(i)\ m_0 = \dfrac{2en}{\lambda} = \dfrac{0{\cdot}6\times1{\cdot}50}{6{\cdot}00\times10^{-5}} = 15{,}000$

---

* The fringes are therefore identical with Newton's Rings (Fig. 12·7) in *appearance*, although *not in origin.* In any case, here the order m is smaller for outer rings, whereas in the case of Newton s rings it is larger for outer rings.

$$\text{(ii)}\quad m_0-10=\frac{2en}{\lambda}\left(1-\frac{\theta^2}{2}\right)=15000\left(1-\frac{\theta^2}{2}\right)$$

$$\therefore\qquad \frac{\theta^2}{2}=\frac{10}{15000}\;\Rightarrow\;\theta=0{\cdot}037\text{ rad.}$$

12·6. Michelson's Interferometer. Michelson devised an instrument in which two *and just two* coherent wavefronts travelling along mutually perpendicular directions were created and later mixed together to show interference. Some of the most remarkable precision results in all physics were obtained from this instrument.

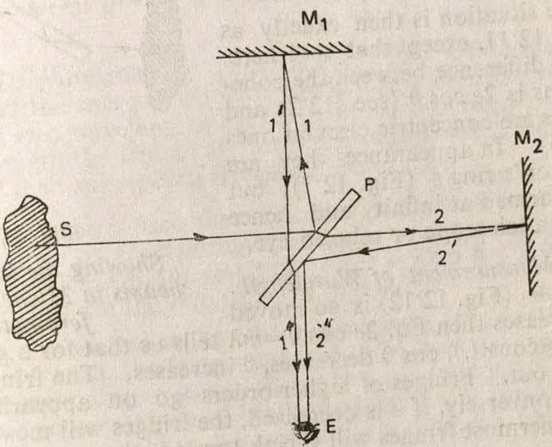


Fig. 12·12

Michelson Interferometer

The essential arrangement (Fig. 12·12) consists of two front-silvered mirrors M_1 and M_2, mutually $\perp^r$, and a glass plate P lightly silvered on the side facing M_2. Tilt of M_2 is adjustable by means of screws (not shown). Mirror M_1 is able to travel normal to itself by means of an accurate screw, the movement being measurable upto (say) 10^{-5} *cm*, *i.e.*, of the order of $\lambda/5$ in the visible range.

Light from source S splits at plate P into beams 1 and 2. These beams are reflected as $1'$ and $2'$ respectively, and parts of these reach the observer's eye E (or a telescope) as $1''$ and $2''$. Interference occurs between these coherent beams.

The simplest way to visualise the functioning is to consider the diagram 12·12 folded about the back-face of plate P. This gives the image of M_2 in P as M_2' as shown in Fig. 12·13. The two interfering beams are then, *in effect*, those that are reflected from M_1 and M_2'. We thus have a parallel air 'film' between M_1 and M_2', with both faces fully reflecting.

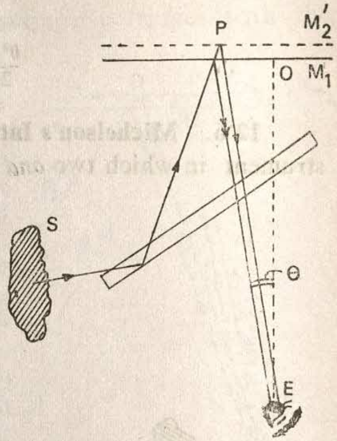


Fig. 12·13

Showing the interfering beams in Michelson Interferometer

The situation is then exactly as with Fig. 12·11, except that $n=1$ here. The path difference between the coherent beams is $2e \cos \theta$ (see §12·2) and the fringes are concentric circular ones (see §12·5). In appearance they are like Newton's ring's (Fig. 12·7), but they are formed at infinity and hence seen with a telescope or relaxed eye.

(*a*) *Measurement of Wavelength.* If mirror M_1 (Fig. 12·12) is so moved that e increases then Eq. $2e \cos \theta = m\lambda$ tells us that for a given order fringe (m=constt.), $\cos \theta$ decreases, θ increases. The fringes therefore move out. Fringes of higher orders go on appearing at the centre. Conversely, if e is decreased, the fringes will move inwards and the innermost fringes will shrink towards the centre and vanish.

In measuring the wavelength of a given light we go on *counting* the fringes q which appear or disappear at the centre of field of view as M_1 is moved by a measured distance $\triangle e$. For the two views *at the centre*, we have

$$2e = m\lambda, \text{ and } 2(e \pm \triangle e) = (m \pm q)\lambda$$

whence

$$\lambda = \frac{2 \triangle e}{q} \qquad \ldots(12{\cdot}21)$$

Notice that m may be an integer or half-integer or anything—in other words, we may make the counts on bright fringes or dark fringes or on any shade. Notice also that absolute value of e is not involved, only *change* of e is involved in eq. (12·21).

Example 12·10. The initial and final readings of a Michelson Interferometer screw are 10·7347 mm and 10·7051 mm as 100 fringes pass. Calculate the wavelength of light used.

Solution. $\triangle e = 10{\cdot}7347 - 10{\cdot}7051 = {\cdot}0296$ mm.

$$\lambda = \frac{2 \times {\cdot}0296}{100} \text{mm.} = 5{\cdot}92 \times 10^{-5} \text{ cm.} \qquad ⊙$$

Example 12·11. An air-cell of thickness 3·00 cm is introduced in one path of a Michelson interferometer, and fringes are obtained with mercury green light ($\lambda = 5{\cdot}46 \times 10^{-5}$ cm). Calculate the number of fringes that pass the field of view

as the pressure in the cell is changed from 76·0 cm to 16·4 cm [n of air at 76·0 cm pressure is 1·000293 and $(n-1)$ is proportional to pressure.]

Solution. $(n-1)$ at 16·4 cm pressure $= \cdot 000293 \times \frac{16\cdot 4}{76\cdot 0} = \cdot 000063$

$\therefore$ δn in changing pressure from 76·0 to 16·4 cm $= \cdot 000230$

$\therefore$ change in one way path $= t.\delta n = 3 \times \cdot 000230 = \cdot 000690$ cm

$\therefore$ number of fringes passing $= \frac{2 \times \cdot 000690}{5\cdot 46 \times 10^{-5}} = 25\cdot 3$

[The converse of this is used for measuring the refractive index of gases.]

(*b*) *Determination of Difference of Two Close Wavelengths.* Imagine that the source used with Michelson's interferometer gives out light of two wavelengths λ_1 and λ_2 whose difference is small. For each wavelength the fringes are formed independently, and we have to consider then superposed fringe-systems.

If the e value is such that the centre has a maximum for λ_1 and also for λ_2, then on superposing the fringes we get very good contrast of intensities, *i.e.*, the resultant fringe system has good *visibility*. The fringes are the said to be *in consonance*.

If the e value makes the centre maximum for λ_1 and minimum for λ_2 or vice versa, the fringes are said to be *in dissonance*, since the fringes corresponding to λ_1 and λ_2 are out of step by half-fringe and the result is poor contrast of bright and dark. In fact if intensities for λ_1 and λ_2 are equal, we get just uniform illumination in the field of view at dissonance.

If the interferometer thickness e is varied by moving mirror M_1 (Fig. 12·12) a *consonance* occurs whenever

$$\frac{2e}{\lambda_1} - \frac{2e}{\lambda_2} = \text{an integer}$$

If the *next consonance* occurs at thickness $e \pm \triangle e$, we have

$$2\triangle e\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) = 1 \qquad \ldots(12\cdot 22)$$

Or, putting $\lambda_2 - \lambda_1 = \triangle\lambda$ and $\lambda_1\lambda_2 = \lambda^2$ we get

$$\triangle\lambda = \frac{\lambda^2}{2.\triangle e} \qquad (\lambda = \text{mean wavelength}) \qquad \ldots(12\cdot 23)$$

For two successive *dissonances* also we get the same result. Therefore, by determining the distance $\triangle e$ by which M_1 has to be moved to pass from one consonance to next (or from one dissonance to next) we get $\triangle\lambda$.

(*c*) *White Light Fringes in Michelson Interferometer.* Let Mirror M_2 be tilted so that the surfaces M_1 and M_2' in Fig. 12·12 are not parallel but inclined, as shown in Fig. 12·14 a. Fringes of equal thickness, located in the film, will now be observed. They will be *straight fringes*. If M_2' and M_1 intersect[*] on a line at O, then the zero order bright fringe will pass through O.

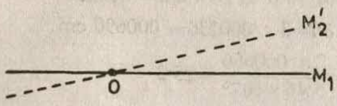


Fig. 12·14

The effective air 'film' for white light fringes

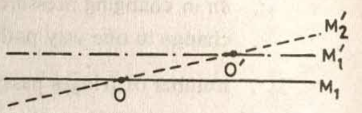


Fig. 12·15

Effect of introducing a sheet in one path

With white light, the zero order fringe will be white but other fringes will be coloured. *Thus the zero-order fringe can be distinguished from the rest.* This is the prime inportance of white fringes.

Imagine now a uniform sheet of thickness d and refractive index n introduced in one of the paths. This introduces an *extra* oneway path $(n-1)\,d$ for this beam and is equivalent to shifting mirror M_1 to M_1' (Fig. 12·15) by a distance $(n-1)\,d$. The observed result is that the zero order fringe shifts from O to O'. If the zero-order fringe is brought back to O by moving the mirror M_1 forward by an amount $\triangle e$ we get

$$(n-1)d=\triangle e \qquad \ldots(12{\cdot}24\ a)$$

Alternatively, if the shift OO' is y and fringe-width with monochromatic light of wavelength λ is w, we have :

$$2(n-1)d=(y/w)\lambda \qquad \ldots(12{\cdot}24\ b)$$

In either case, if n is known, we get d. This is the use made inpractice for sheets of mica, cover glass, etc. But more commonly it is used for determining n of gases, d being known.

Example 12·12. In a Michelson interferometer set for white light a cover-glass of thickness $2{\cdot}31\times10^{-3}$ cm and $n=1{\cdot}53$ is introduced in one path. If fringe-width with *Hg* green light ($\lambda=5461$ A) is $1{\cdot}82\times10^{-2}$ cm, by what distance would the white fringe shift ?

Solution. Additional optical path$=2(1{\cdot}53-1)\times2{\cdot}31\times10^{-3}$ cm

$$=2{\cdot}45\times10^{-3}\text{ cm.}$$

Equating this with $[y/(1{\cdot}82\times10^{-2})]\times5461\times10^{-8}$ cm gives

$$\text{shift } y=\frac{2{\cdot}45\times10^{-3}\times1{\cdot}82\times10^{-2}}{5{\cdot}491\times10^{-5}}=\mathbf{0{\cdot}82\ cm}$$

[Note that use of Eq. (12·24 a) would need measurement $\triangle e$ which is$\sim10^{-3}$ cm, whereas use of Eq. (12·24 b) needs measurement of y which is$\sim$1 cm—in effect $\sim10^3$ fold magnification.]

⊙

---

[*] In Michelson interferometer this is possible because M_2' is an image, not the actual mirror.

(*d*) *Standardization of the Metre.* The standard metre (distance at 0°C between the fine lines engraved on a platinum-iridium rod preserved in the Bureau of Weights and Measures at Paris) was measured by Michelson with this interferometer in terms of *the red line of cadmium*, in the years 1892-95, giving $\lambda_R = 6438{\cdot}472$ A, which means an accuracy of 1 in $\sim 10^7$. We shall omit the details of the method. Although more accurate determinations were later made using Fabry-Perot interferometer and employing more monochromatic spectral lines, the historical importance of this work stands.

(*e*) *Determining earth's motion through ether.* The original design of the Michelson interferometer was intended to see if the velocity of light along the earth's motion in space and along the perpendicular to it differed—a test for the existence of absolute space (see Chapter 24). The interferometer was placed on a rock about 1 m^3 size, which was floated on mercury. One arm was set parallel to earth's motion in space and interference fringes were obtained. On rotating the instrument by a right angle, a small fringe-shift was expected, if there existed an absolute space. The experiments of Michelson and Morley (1881 to 1887) showed a null result ! And this shook the foundations of much philosophical thought and in the hands of Einstein gave birth to the Theory of Relativity. (See Chapter 24).

12·7. The Twyman and Green Interferometer. On the technological side, one of the modifications of Michelson's interferometer widely used is the Twyman-Green interferometer (Fig. 12·16). A pinhole S with collimating lens L_1 is used as the source, and a focussing

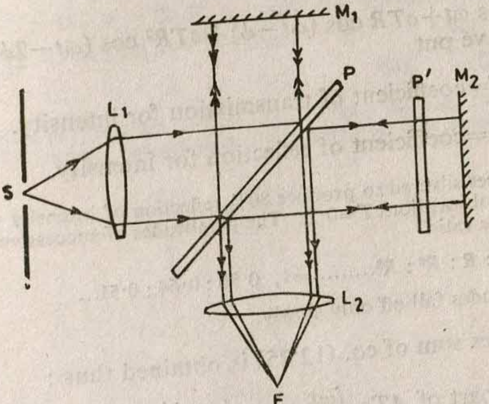


Fig. 12·16

The Twyman-Green Interferometer

lens L_2 with eye placed at its focus for viewing. With M_2' (Fig. 12·16) set parallel to M_1 the whole surface of M_1 shows a uniform intensity. Now, a test plate P' is introduced. If it is uniform in thickness and structure, the field of view remains of uniform intensity. But if there are irregularities in P', fringes show up around those irregularities and one can correct them (by polishing etc.) while viewing the improvements. If P' is a convex lens, then M_2 is replaced by a convex mirror of radius of curvature equal to the focal length of the lens. If mirrors are to be tested and corrected, P' is an ideal lens and M_2 is the mirror under test. In effect the arrangement is used to make the test plates, lenses and mirrors accurate to better than $\lambda/10$ or so in precision.

12·8. Multiple Beam Interferometry. In § 12·1 (Fig. 12·1) we showed how in multiple reflection at ordinary surfaces only the first *two* beams are of appreciable amplitudes. The situation changes altogether if the *surfaces of the film are partially silvered.* Then all the beams become of comparable amplitude and the intensity distribution in the interference pattern is very different from the $\cos^2(\delta/2)$ type. This forms the basis of 'multiple beam interferometry'.

In Fig. 12·1 let the amplitude of incident wave be a, and let the coefficients of transmission and of reflection *for amplitude* be t and r respectively. Let the phase term of the first wave be $\cos \omega t$, and successive waves have phase retardation ϕ. Then the resultant disturbance when all the transmitted beams are collected to a focus and interfere is given by

$$y=aT\cos\omega t+aTR\cos(\omega t-\phi)+aTR^2\cos(\omega t-2\phi)+\ \ldots(12{\cdot}25)$$

where we have put

$t^2 \equiv T$=coefficient of transmission for intensity.

$r^2 \equiv R$=coefficient of reflection for intensity.

With faces silvered to produce 80% reflection of intensity we have $R=0{\cdot}8$ and (neglecting absorption) $T=0{\cdot}2$. The amplitudes of successive terms in eq. (12·25) are in the ratio

$$1:R:R^2:R^3.........=1,\ 0{\cdot}80:0{\cdot}64:0{\cdot}51\ldots$$

Thus the amplitudes fall off only slowly.

The series sum of eq. (12·25) is obtained thus :

$$y=\text{real part of } ATe^{i\omega t}(1+Re^{-i\phi}+R^2e^{-2i\phi}+\ldots)$$

$$=\text{real part of } ATe^{i\omega t}(1-Re^{-i\phi})^{-1}=\frac{AT}{p}\cos(\omega t-\alpha)\ \ \ldots(12{\cdot}26)$$

where

$$p=\sqrt{1+R^2-2R\cos\phi}\ \ \ldots(12{\cdot}27)$$

and α is presently of no interest to us.

If I_0 is the maximum intensity (for $\phi=\phi_m=2n\pi$), the intensity I for a plase difference $\phi m+\delta$ is given by

$$\frac{I}{I_0}=\frac{(1-R)^2}{I+R^2-2R\cos\delta}=\frac{1}{1+F\sin^2(\delta/2)} \qquad ...(12{\cdot}28)$$

where F depends only on reflectivity R, and has the value

$$F=\frac{4R}{(1-R)^2} \qquad ...(12{\cdot}29)$$

Thus the intensity variation is not of the $\cos^2(\delta/2)$ type, as in unsilvered films or Michelon interferometer. Since F has a large value, the intensity falls rapidly as ϕ deviates from $2n\pi$. The half width of the maxima in terms of phase angle (δ_h) is given by setting $F\sin^2(\delta_h/2)=1$. For large F we get

$$\delta_h=\frac{2}{\sqrt{F}} \qquad ...(12{\cdot}30)$$

F called *coefficient de finesse* (coefficient of fineness) of the system. For R close to 1, F is quite large and hence the maxima are very sharp.

Example 12·13. Obtain a plot of I/I_0 against δ when $R=0{\cdot}80$.

Solution. First from eq (12·29), $F=80$ (check it)

Then for various of δ between 0 and ±180° we calculate I/I_0 as follows from eq. (12·28) :

δ	0	±10°	±30°	±60°	±120°	±180°
$\sin^2\delta/2$	0	·0076	·0670	·2500	·5868	1·000
$1+F\sin^2\delta/2$	1	1·61	6·36	21	48	81
I/I_0	1	·62	·16	·048	·021	·012

The graphical plot is given in Fig· 12·17, where we also show for comparison the $\cos^2\delta/2$ curve. The contrast in sharpness is to be noted. ⊙

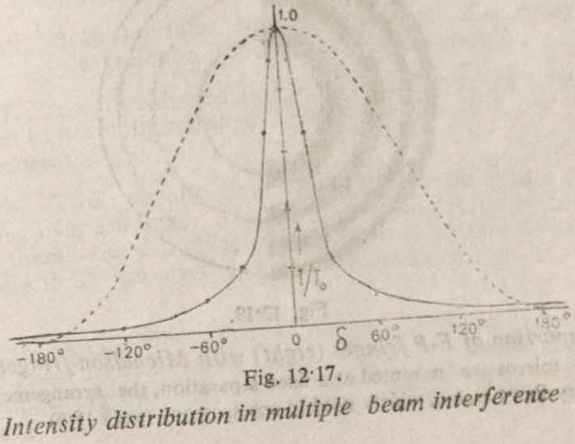


Fig. 12·17.

Intensity distribution in multiple beam interference

Due to sharpness of the maxima multiple beam interferometry has been widely employed wherever interference fringes are used to measure small thicknesses or changes therein : for example, in measuring Young's modulus, thermal expansion, etc., and studying crystal growth.

We shall, however, discuss just one case of some importance in spectroscopy—the Fabry-Perot interferometer.

12·9. Fabry and Perot Interferometer. This consists of two glass plates P_1, P_2 (Fig. 12·18) thinly silvered or aluminsed to reflect 80 to 90% intensity, the silvered surfaces being put face to face and parallel. One of the plates can be moved perpendicular to its own plane by means of an accurate screw.*

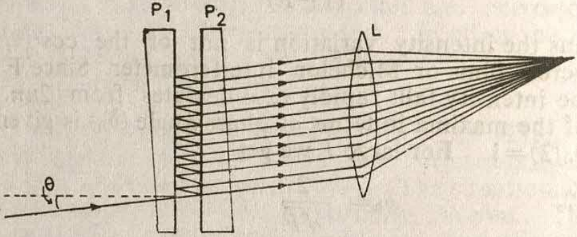


Fig. 12·18.

Working of a Fabry-Perot Interferomenter

If a wide source of monochromatic light is viewed through the system, circular fringes are seen as in Michelson's interferometer. Now if the movable plate is shifted by a distance $\triangle e$ we find a number of fringes emerging (or collapsing), as in the case of Michelson interferometer (§ 12·6 *a*) and eq. (12·21) leads to determination of λ.

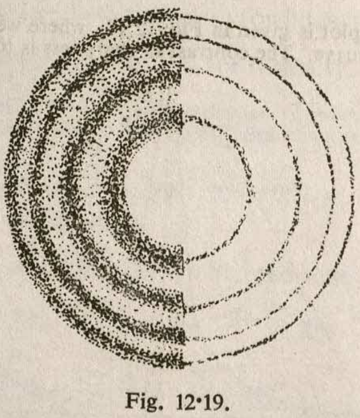


Fig. 12·19.

Comparison of F-P fringes (right) with Michelson fringes (left)

---

* If the mirros are mounted at a fixed separation, the arrangement is called a Fabry-Perot *etalon*, widely used in spectroscopy (see § 15·6)

However, the nature of fringes in Fabry-Perot interferometer is very different, as shown by eq. (12·28) and Fig. (12·17). To emphasise we schematically show in Fig. 12·19 the view of Fabry-Perot fringes alongside the Michelson fringes for the same separation e.

PROBLEMS

12·1. Interference fringes in thin films are quite distinct when seen in reflection, but very indistinct in transmission. Explain. Will the statement hold good for films with reflecting coatings ?

12·2. Explain why you need a *wide source* for seeing interference in films. Also explain why the observing instrument has to be focussed in a particular plane, unlike Young's fringes (Chapter 11) which are seen over a wide region of space.

12·3. A thin soap film seen in sodium yellow light by normal reflection appears totally dark. Deduce the possible values of film thickness. ($n=1·33, \lambda=5893\,A°$).

12·4. An interferometric fringe method is used to measure changes of position. If one can measure shifts upto one-fifth of a fringe, deduce the least count of the device (λ used is $5·89\times10^{-5}$cm).

12·5. Non-reflecting coatings are $\lambda/2n$ thick. Explain. Could they be (say) $7\lambda/2n$ thick ? (n is refractive index).

12·6. Explain the following :

(*a*) An infinitely thin film seen in reflected light should appear dark.

(*b*) A thick film seen in white light shows no colours.

(*c*) In thin films successive fringes of a given colour represent a variation of thickness$\sim\lambda/2n$.

12·7. A wedge-shaped air film with wedge angle 1 minute of arc is seen in normally reflected light $\lambda=5·46\times10^{-5}$ cm. Deduce the fringe width.

12·8. Two optically plane glass plates, each 6 cm long, are placed one over another. At one end a mica strip is introduced between them. Seen in reflected light of $\lambda=5·89\times10^{-5}$ cm we find 100 fringes in 3·0 cm length. Compute the thickness of the mica strip.

12·9. An air film is formed between the upper face of a crystal and lower face of a glass plate held by a separate support. A microscope cross-wire is focussed over a bright fringe using $\lambda=5·46\times10^{-5}$ cm. As the crystal is heated from 20°C to 80°C, 56 fringes pass across the cross-wire. Crystal height being 2·5 cm, deduce its coefficient of linear expansion.

12·10. A man standing 2·0 m above an oil film over water observes greenish colour at a point 2·5 m from his feet. Deduce the possible thickness of the oil film there. (λ_{green}=5500 A^0, n_{oil}=1·40, n_{water}=1·33).

12·11. In an experiment to observe Newton's rings we may find that (*i*) $D_{m+r}^2-D_m^2$ does not come out as independent of *m*, (*ii*) the rings are not quite circular, (*iii*) the centre is not dark, but is bright. What would you conclude in each case ?

12·12. Newton's rings are observed between a spherical surface of radius of curvature 120 cm and a plane plate. The diameters of 5th and 15th bright rings are 0·314 cm and 0·584 cm. Calculate the diameters of 25th and 37th bright rings, and also the wavelength of light used.

12·13. In a Newton's rings arrangement $D_{m+r}^2-D_m^2$ came to 0·204 cm^2 with air film. Deduce the $D_{p+r}^2-D_p^2$ value (*r* same as before) if a drop of liquid of n=1·473 is introduced as the film. Will the fringes (seen in reflection, as usual) be more intense with the liquid as film ?

12·14. (*a*) In what respects are Haidinger's fringes distinct in character from fringes in thin films ?

(*b*) Fringes in Michelon's interferometer and Fabry and Perot interferometer do not differ in character, but differ widely in intensity distribution. Comment.

12·15. In a Michelson interferometer set for circular fringes, 268 fringes appear as the movable mirror shifts by 0·0774 mm. Deduce λ of the light used.

12·16. In a Michelson inteferometer circular fringes are observed with sodium yellow light (λ_1=5890*A*, λ_2=5896*A*). Describe and explain what changes are observed as the 'separation' of mirrors is increased from 0·10 mm to 1·40 mm.

12·17. A Michelson interferometer, set for straight fringes, shows 15 fringes over a width of 1·90 cm using sodium light (λ= 5893 *A*). Calculate (i) the angle of inclination between the 'mirrors', (ii) the number of fringes over the same width with mercury violet (λ=4358*A*), and (iii) the number of fringes which will cross an observation point as the movable mirror is shifted by 0·0374 mm using sodium light.

12·18. A gas-cell of 10·0 cm thickness is introduced in one path of a Michelson interferometer and white light fringes are obtained with gas pressure 100 cm of mercury. The gas pressure is gradually reduced to 50 cm of mercury. Describe what would be observed. To get the central fringes back at the original place one mirror needs shift of $1{\cdot}9 \times 10^{-3}$ cm. Calculate n of the gas at 76 cm pressure.

12·19. Compute the half-width of the Fabry-Perot fringes for $R=0{\cdot}80$, 0·90, 0·95 and 0·99, using the theory of § 12·8. Comment on the validity of the theory for R approaching too near unity.

12·20. How would the view of fringes in F-P interferometer at 'consonance' and 'dissonance' (see § 12·6 *b*) be different from that in a Michelson's interferometer ?

12·21. For a truly monochromatic single-line source Michelson interferometer fringes should have $\cos^2(\delta/2)$ intensity distribution, giving visibility $V=1$ (Eq. 12·16). But every spectral line has a finite half-intensity width $\Delta\lambda_{1/2}$. How will it affect V as the path differenece p increases ?

13

Diffraction of Light (I)

13·1. Diffraction. *We may broadly define diffraction as departures from the expectations of ray optics* when a beam of light passes through an aperture or by an obstacle. Diffraction effects are dominantly observed when the aperture or obstacle has dimensions comparable with the wavelength of light. Experimentally, all waves exhibit diffraction. One demonstration can be made with waves on water using the ripple tank apparatus (of § 10·8). A schematic represention of the view is given in Fig. 13·1, in which plane waves generated by a dipper *AA* pass through an aperture *CD* which has a width a few times λ. On the right side of the aperture the ripples are not limited to the direction *P*, but spread out. The ripples fade out as directions *Q,Q* are reached, and reappear (though weakly) along directions *R,R* and so on. What happens on changing the width of the aperture *CD* is to be better seen than described.

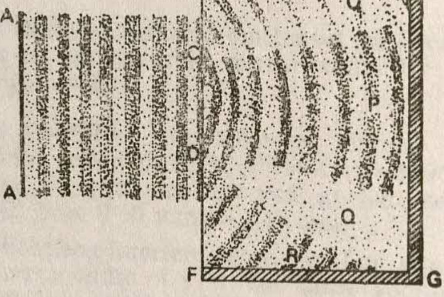


Fig. 13·1.

Ripple tank demonstration of diffraction at an opening CD

With visible light the following demonstrations are very educative :

(*a*) Scratch a sharp line on a blackened glass plate and through it view a filament lamp a few metres away. You will see the filament rather spread out and then a series of several filaments on either side with successively decreasing intensity. See what happens if the plate is roated in its own plane.

(*b*) Look at a distant street lamp through a fine handkerchief or a saree curtain. Instead of a single lamp you will see a whole network of lamps arranged on mutually perpendicular axes. If you rotates the handkerchief, the pattern rotate too. The finer the weaving of the cloth, the larger is the separation of successive 'images'.

(*c*) Hold in front of your eye a glass plate sprinkled with some graded lycopodium powder and look at a hole (a few mm dia) in the cover of a mercury discharge lamp placed a few metres away. The hole appears as a spread out disc, surrounded by a few alternately dark and bright rings. Using yellow, green and blue filters in turns see how the angular diameters of the rings vary with colour.

In Fig. 13·2 we have *S* as a source point, *AB* an aperture, and *SAC* and *SBD* the projections within which we ordinarily expect light to be limited, if ray optics is followed. But if we apply Huygens principle, the secondary wavelets (as shown) should reach outside the limits *AC* and *BD* not marginally but to a considerable extent. But we know that diffraction effects in light occur only marginally. Are we then to give up the Huygens' principle, or the wave theory itself ? Fresnel in 1815 showed by careful analysis that (*i*) Huygens' wavelets cancel out for regions far outside *AC* and *BD* (*ii*) for regions well within *AC* and *BD* the Huygens' wavelets add up to give the same intensity as in the absence of the aperture, and (*iii*) for regions close to the limits *AC* and *BD* (both inside and outside) the wave theory predicts departures of intensity which match the experimental observations. In essence the theory involves *summing up the contri-*

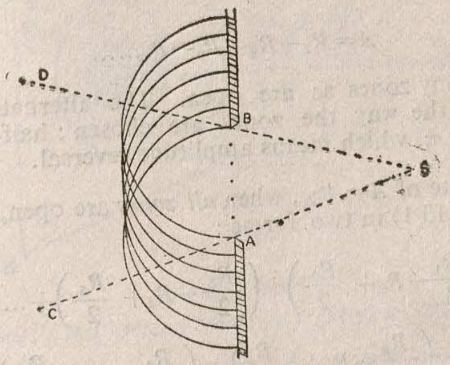


Fig. 13·2.

Explaining the difficulties with simple Huygens' Principle *butions from different parts of the available wavefront, taking care of their phase relations.*

One can argue that such 'summing up' is done to explain interference patterns also. The only difference is that in the diffraction case we have one continuous wavefront over which the effect is to be integrated, whereas in the interference case we have a discrete set of waves whose effect is to be summed up. So the differnce is between summing up a series and integrating over some given limits. Some authors treat this distinction as trivial and use the word *interference* to cover all cases of superpositions of coherent waves. But we shall use the term *diffraction* as distinct from *interference* as a convenient usage.

13·2. Fresnel's Half-period Zone Method.* Let a wave front be limited in width by an aperture. To deducce the intensity at any forward point (P), Fresnel devised the *half-period zone* method. The steps are as follows :

(a) We divide the wavefront into zones such that the waves reaching P from extreme edges of a given zone have a phase difference of half-period. Each zone is then called a *h.p.* zone. The waves from corresponding parts of successive *h.p.* zones will reach P with phase difference π.

(b) We consider the amplitude of disturbance (R_n) from the n^{th} zone as proportional to (*i*) the area of the zone, (*ii*) cosine of the in clination and (*iii*) reciprocal of the distance to P. The last two have only a small effect.

(c) We find the amplitude A of the resulant disturbance at P by adding up

$$A=R_1-R_2+R_3-R_4\ldots\ldots \qquad \ldots(13{\cdot}1)$$

including as many zones as are open. The alternate + and − signs here arise from the way the zones are chosen : half-period means phase difference π, which means amplitude reversal.

An estimate of $A=A_\infty$, when *all zones* are open, can be had by writing Eq. (13·1) in two forms :

$$A_\infty=\frac{R_1}{2}+\left(\frac{R_1}{2}-R_2+\frac{R_3}{2}\right)+\left(\frac{R_3}{2}-R_4+\frac{R_5}{2}\right)+\ldots\ldots\text{to }\infty$$

$$A_\infty=R_1-\frac{R_2}{2}+\left(\frac{R_2}{2}-R_3+\frac{R_4}{2}\right)-\left(\frac{R_4}{2}-R_5+\frac{R_6}{2}\right)-\ldots\ldots\text{to }\infty$$

If the bracketted terms are all positive or all negative, we see that A_∞ lies between $\frac{1}{2}R_1$ and $R_1-\frac{1}{2}R_2$. Hence as a near approximation

$$A_\infty\approx\tfrac{1}{2}R_1 \qquad \ldots(13{\cdot}2a)$$

---

* See also Problems 13·13 to 13·15

If zones numbered 0 to n are closed and only those numbered $n+1$ to ∞ are open, we get

$$A_n' \approx \pm \tfrac{1}{2} R_{n+1} \qquad \text{...(13·2 b)}$$

and if the zones numbered 1 to n only are open, we get

$$A_n \approx \tfrac{1}{2} R_1 \pm \tfrac{1}{2} R_{n+1} \qquad \text{...(13·2 c)}$$

where the upper sign applies for n odd and the lower sign for n even.

In case the apertures permit fractions of some *h.p* zones to contribute, then in eq. (13·1) we must take the terms R multiplied by the appropriate fractions.

It will be seen that the Huygens' wavelets reaching P from within a *h.p.* zone are not all in the same phase—the extreme difference of phase being π. An approximation is therefore involved in taking the R_n values proportional to zone area, as if the entire *h.p.* zone contributes in one phase. But this is not a serious limitation for two reasons. Firstly, the phase difference for contributions from different parts within a zone will introduce a correcting factor which will be very nearly the same for all the R values. Secondly, the wavelets from *corresponding parts* of neighbouring zones have rigorously the phase difference π in reaching P.

13·3. Fresnel's Diffraction at a Circular Aperture or Disc.

(a) Half-period zones. In Fig. 13·3 let the point source be at a very large distance on the left so that the wavefront $ABCD$ is plane.

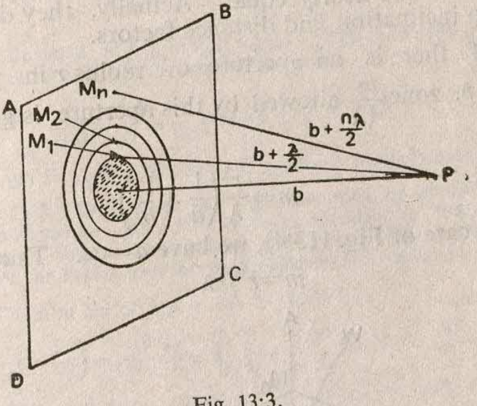


Fig. 13·3.

Fresnel's half-period zones on a plane wavefront

For an observation point P, we drop PO normal to the wavefront. If $PO=b$, we draw circles on the wavefront with radii OM_1, OM_2... OM_n, where M_n is so chosen that

$$PM_n - b = \frac{n\lambda}{2} \qquad \text{...(13·3)}$$

These circles divide the wavefront into *h.p.* zones, because secondary wavelets from successive zones have to travel an extra distance $\lambda/2$ in reaching P.

If the source is at a finite distance a from the aperture, the *h.p.* zones may be drawn as shown in Fig. 13·4. The wavefront WOW is now spherical, and we may choose $M_1, M_2 \ldots M_n$ on this surface; alternatively, we may use plane AOB as the projection of the wavefront and choose M_n on that, with

$$SM_n + M_nP - (a+b) = n\,\frac{\lambda}{2} \qquad \ldots(13{\cdot}4)$$

The radius $r_n = OM_n$ is then given by

$$\frac{1}{2}\frac{r^2{}_n}{a} + \frac{1}{2}\frac{r^2{}_n}{b} = \frac{n\lambda}{2} \qquad \ldots(13{\cdot}5)$$

using binomial theorem since $r_n \leqslant a$ or b. That gives

$$\left(\frac{1}{a}+\frac{1}{b}\right) r^2{}_n = n\lambda \qquad \ldots(13.6)$$

Certain geometrical facts are worth noting :

(*i*) r_n is proportional to $\sqrt{n}$, so that successive *h.p.* zones have narrower spacing (See Fig. 13·6).

(*ii*) Area πr_n^2 of n *h.p.* zones is proportional to n, which means that successive *h.p.* zones have equal areas. Hence amplitudes $R_1, R_2, \ldots R_n \ldots$ of waves contributed by successive *h.p.* zones to the disturbance at P are nearly equal. Actually, they decrease slowly with n due to inclination and distance factors.

(*iii*) If there is an aperture of radius r in plane AB, the number of *h.p.* zones $\frac{m}{(\lambda}$ allowed by this aperture is given, from Eq. 13·6), by

$$m = \frac{r^2}{\lambda}\left(\frac{1}{a}+\frac{1}{b}\right) \qquad \ldots(13{\cdot}7)$$

For the case of Fig. (13·4), we have $a = \infty$. That leads to

$$m = r^2/b\lambda \qquad \ldots(13{\cdot}8)$$

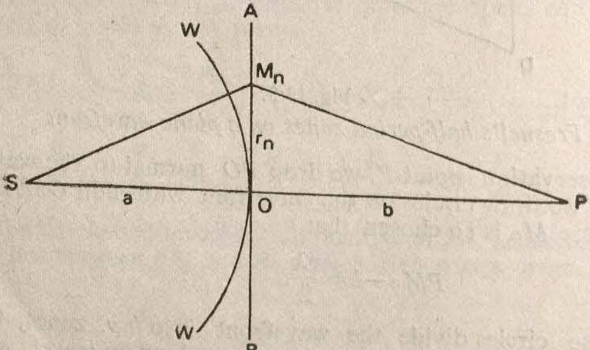


Fig. 13·4.

(*b*) **Axial observation points.** Now we can examine the intensity at an axial observation point for a given set of r, a and b. As Eq. (13·2 *c*) would show, the disturbance at P is strong when m is odd and weak (nearly zero) when m is even. Hence the observed axial point becomes alternately bright and dark according as variation of either b or a makes m odd or even, respectively.

Example 13·1. A plane wavefront of light of wavelength $5{\cdot}5\times10^{-5}$ cm is intercepted by a circular aperture of 5·0 mm. diameter. Calculate the number of half-period zones in the aperture for an axial observation point distant (i) 1·0 metre, (ii) 5·0 metres from the aperture.

Solution. For case (i),

$$m=\frac{(0{\cdot}25\text{ cm})^2}{(100\text{ cm})\,(5{\cdot}5\times10^{-5}\text{ cm})}$$
$$=11.4\ h.p.\text{ zones}$$

For case (ii) b is 5 times larger. Hence

$$m=\frac{11{\cdot}4}{5}=2.3\text{ h.p. zones.}$$

⊙

Example 13·2. Along the axis of a circular aperture of diameter 2·00 mm. on one side is a point source at 1·00 metre distance emitting light of $\lambda=5{\cdot}46\times10^{-5}$ cm, and on the other side is a point of observation at a varying distance D. For what values of D does the aperture involve (i) two, (ii) four, (iii) six half-period zones ?

Solution. For use in Eq. (13·7) we are given $r=0{\cdot}100$ cm, $a=100$ cm, $\lambda=5{\cdot}46\times10^{-5}$ cm and $b=D$, which varies as n varies.

$$\therefore\quad \frac{1}{100}+\frac{1}{D}=\frac{m\lambda}{r^2}=m\times5{\cdot}46\times10^{-3}\text{ cm}^{-1}$$

Form $=2, 4, 6$ we get $D=\mathbf{1{\cdot}09\times10^3,\ 84{\cdot}4,\ 43{\cdot}8}$ cm respectively (check). ⊙

Example 13·3. On the axis of a circular aperture of radius 2·0 mm, on one side lies a point source of $\lambda=5{\cdot}5\times10^{-5}$ cm at distance 2·0 metres. On the other side a point of observation is moved along the axis from infinity to closer distances. Deduce the first 3 position where minima are observed,

Solution. The number of *h.p.* zones is given by

$$m=\frac{r^2}{\lambda}\left(\frac{1}{a}+\frac{1}{b}\right)=\frac{(0{\cdot}2)^2}{5{\cdot}5\times10^{-5}}\left[\frac{1}{200}+\frac{1}{b}\right]=3{\cdot}64+\frac{727}{b}$$

Minima occur when m is even. Starting with $b=\infty$, the first three stages are given by $m=4, 6, 8$. Hence for the three cases

$$\frac{727}{b}=0{\cdot}36\qquad\frac{727}{b}=2{\cdot}36,\qquad\frac{727}{b}=4{\cdot}36$$

⊙

$\therefore\ b=$**2020 cm, 308 cm, 167 cm.**

(*c*) **Off-axis observation points.**

For off-axis observation points we note the following. A sideways displacement y from p (Fig. 13·4) causes the *h.p.* zone system to move sideways by $ya/(a+b)$. If this equals the width of one outermost zone m, the result is to replace contribution R_m by

$\frac{1}{2}R_m-\frac{1}{2}R_{m+1}\approx 0$. In effect it removes R_m. If $ya/(a+b)$ equals the widths of 2,3,4...outermost *h-p.* zones, the effect is to remove the contributions of 2,3,4...outermost zones. The result is that the intensity alternately rises and falls as P goes sideways, so that by rotating fig. 13·4 about axis SP we may visualise alternate bright and dark rings. These are shown schematically in Fig. 13·5 (*a*) and (*b*) when the axial point is bright and dark, respectively.

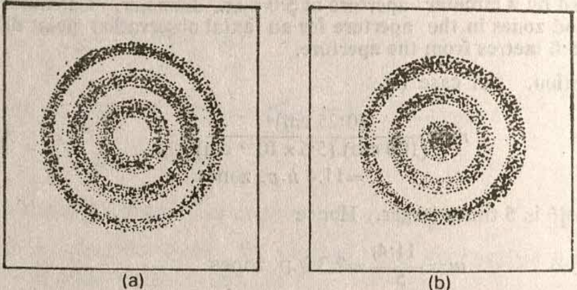


Fig 13·5

Alternate bright and dark rings

(d) Case of circular disc.

For the case of circular disc also we may use Fig. 3·4, replacing the aperture by the disc of radius r. The disc covers up the first m half-period zones and only the zones numbered $m+1$ to ∞ send light to P. So the amplitude at P is given by Eq. (13·2*b*) :

$$A=R_{m+1}-R_{m+2}+\ldots\ldots\text{to}\infty \approx \tfrac{1}{2}R_{m+1}$$

The point P is thus always bright.

As m increases (*i.e.*, either a or b in Fig. 13·4 decreases), R_{m+1} decreases and hence the *intensity continuously decreases*. This is in contrast with the case of a circular aperture, where alternate bright and dark position occur along the axis.

For off-axis positions of P the analysis for a circular *disc* differs from that for the *aperture* in the sense that as $ya/(a+b)$ equals the widths of 1, 2, 3... outermost *covered* zones, the effect is to *add* the contributions of 1, 2, 3... outermost zones. Thus we get concentric bright and dark rings, like those shown in Fig. 13·5 (*a*).

13·4. Zone plate. An interesting application of Fresnel's half-period zones is a zone plate. It consists of a number of alternate transparent and opaque concentric circular zones on a glass plate (Fig.13·6). The radii of *boundaries* of successive zones are given by $r, r\sqrt{2}, r\sqrt{3}, \ldots\ldots r\sqrt{n}\ldots$, where the quantity r is very small, usually fraction of a millimeter.

In Fig. 13·4, let the zone plate be placed in plane AB, with its centre at O. Then, setting $n=1$ in eq. (13·6) we see that the zones on it will become *h.p.* zones if S and P are so chosen that*

---

* The zones of the zone plate are not h.p. zones unless this condition is satisfied.

$$\frac{1}{a}+\frac{1}{b}=\frac{\lambda}{r^2}=\frac{1}{r^2/\lambda} \qquad \ldots(13\cdot9)$$

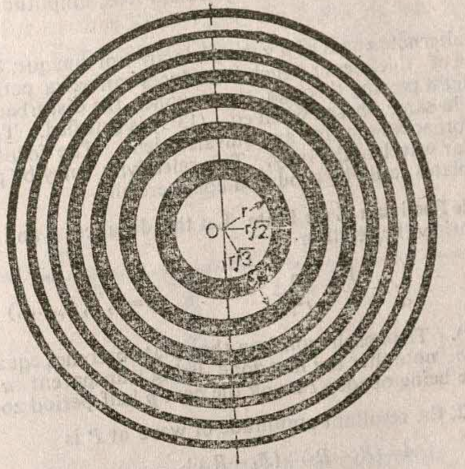


Fig. 13·6.

The zone plate (schematic)

If all the zones were open, resultant amplitude of the disturbance at P would be $\frac{1}{2}R_1$ (see eq. 13·2*a*). But in the zone plate only the alternate zones are kept open. Hence the resultant amplitude at P is given by

$$A=R_1+R_3+R_5+\ldots\ldots$$

These add up to a large value compared with $\frac{1}{2}R_1$. Hence P becomes an intense *image* of S when eq. (13·9) is satisfied.

We thus say that *a zone plate acts like a lens*. What is its focal length ? Eq. (13·9) compared with a lens equation shows :

$$f=r^2/\lambda \qquad \ldots(13\cdot10)$$

where r is radius of the central zone of the zone plate. You may notice that the focal length given by eq. (13·10) depends sharply on the wavelength and *decreases* as the wavelength increases. This in unlike the usual lenses.

Example 13·4. A zone plate is to be constructed with focal length 50 cm for $\lambda=5{\cdot}0\times10^{-5}$ cm. Calculate its first radius. What will be its focal length for $\lambda=4{\cdot}0\times10^{-5}$ cm ?

Solution. From eq. (13·10), $r_1^2=f\lambda=50 \text{ cm}\times5{\cdot}0\times10^{-5}$ cm

$$\therefore \quad r_1=5.0\times10^{-2} \text{ cm}=\mathbf{0{\cdot}50 \text{ mm.}}$$

Again, from eq. (13·10),

$$f_2=f_1.\frac{\lambda_1}{\lambda_2}=50\times\frac{5}{4}=\mathbf{62{\cdot}5} \text{ cm.}$$

Phase Reversal Zone Plate. It may be noted that while a lens collects all the 'disturbances' R_1, R_2.........without phase differences, the zone plate collects light from only the alternate ones. Hence the image is much less intense compared to the image formed by an equivalent lens; amplitude$\sim 1/2$, hence intensity $\sim 1/4$.

If the alternate zones of a zone plate are not opaque, but covered by (say) some gelatine of thickness t which introduces an extra path difference $(\mu-1)t = \lambda/2$ in the light passing through these zones, the disturbances from all zones are collected in same phase when eq. (13·9) is satisfied. The intensity of the image then approaches that for an equivalent lens. The focussing holds, however for a particular wavelength only. This selectivity may be put to use if phase-reversal zone plates could be produced commercially.

Multiple Focii in a Zone Plate. Let the distances a and b (Fig. 13·7) be so chosen as to satisfy the relation

$$\frac{1}{a}+\frac{1}{b}=\frac{p}{r^2/\lambda} \qquad (p=2, 3, 4,......) \qquad ...(13{\cdot}11)$$

instead of (13·9). Then each 'zone' on the plate becomes equal to p half-period zones. We now note that the first zone being transparent *allows* p half-period zones, next zone being opaque *cuts off* the next p half-period zones, and so on.

For $p=2$, the resultant amplitude of wave at P is

$$A=(R_1-R_2)+(R_5-R_6)+......$$

This is nearly zero, since $R_1 \approx R_2$, $R_5 \approx R_6$, etc.

For $p=3$, each zone of the plate equals 3 *h.p.* zones. Hence

$$A=(R_1-R_2+R_3)+(R_7-R_8+R_9)+...$$

(The missing terms are cut off by the opaque zones). Since $R_2 \approx R_3$, $R_8 \approx R_9$... the sum becomes

$$A=R_1+R_7+R_{13}$$

which is large (compared with $\frac{1}{2}R_1$) and produces an image of S at P. Eq. (13·11) with $p=3$, gives the new focal length as

$$f'=r_1^2/3\lambda=f/3$$

There are further focal lengths $f/5$, $f/7$, etc. But there is no image formation corresponding to $f/2$, $f/4$ etc.

It is to be noted that the amplitude of image corresponding to focal length $f/3$ is about one-third of that due to the primary focal length f, hence intensity is about one-ninth.

13·5. Explanation of Rectilinear Propagation. In. Fig. 13·7 let AB be an obstacle between a point source S and a screen, on which P_0 is the grometrical projection of SA. For a general observation point P the pole is at O, around which Fresnel's *h.p.* zones may be drawn. The intensity at P will depend on how many *h-p.* zones are able to contribute.

Let us first deduce an expression for the relation between x and the *h.p*· zone number n which touches A. The condition is

$$(SA+AP)-SP=n\lambda/2$$

and from the geometry and binomial theorem ($x<<a,b$) we get

$$n=\frac{ax^2}{\lambda b(a+b)} \qquad \text{...(13·13)}$$

Typically for $a=b=50$ cm and $\lambda=5000$ A, we find that $x=1$ cm gives $n=200$. This number increses in proportion to x^2.

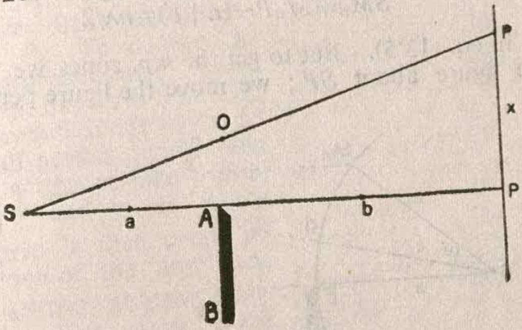


Fig. 13·7

For explaining rectilinear propagation

In the absence of the obstacles, *h.p.* zones upto $n=\infty$ contribute, leading to amplitude $A=\frac{1}{2}R_1$, vide eq. (13·2*a*). Amplitude elsewhere is to be compared with this value. If P is a few mm above P_0, we have n so large that the contribution from the open zones approaches $\frac{1}{2}R_1$. If P is a few mm below P_0, then for a large number of *h.p.* zones n there is no contribution at P and contributions only from zones $n+1$ to ∞ exist—that too in fractions. Thus the total amplitude is a small fraction of $\frac{1}{2}R_{n+1}$. This is $\ll\frac{1}{2}R_1$, and hence the intensity at P is effectively zero.

Thus, rectilinear propagation is explained on the wave theory. But the theory also says that for P near the edge P_0—whether above or below—there will be departures from the expectations from rectilinear progation.

To learn about the nature of the departures, let the amplitude of disturbance at P due to the *n fully open h.p.* zones be R_a and that from the rest of the *partially open h.p.* zones be R_b. For P above P_0 each of R_a and R_b will vary with x. In particular, n becoming alternately even and odd will make R_a alternately small and large (see eq. 13·2*c*). Thus intensity variations (*bright and dark bands*) will be observed.

For P below P_0 (inside the shadow), the contribution R_a is absent, and R_b falls monotonously as n increases. So the *intensity falls continuously* as we go deeper into the shadow.

13·6. Diffraction at a Straight Edge. (*a*) **Half period zones.** In fig. 13·8 let S be a slit and E a straight edge, both perpendicular to the plane of paper. Let P be an observation point, which gives O as the pole with respect to which the *h.p.* zones are to be drawn. The wavefront EOM_n is now cylindrical with S as axis. On it we choose $M_1, M_2 \ldots M_n$ such that

$$SM_n + M_nP - (a+b) = n\lambda/2$$

exactly as in eq. (13·5). But to get the *h.p.* zones we now do not rotate the figure about SP ; we move the figure perpendicular to

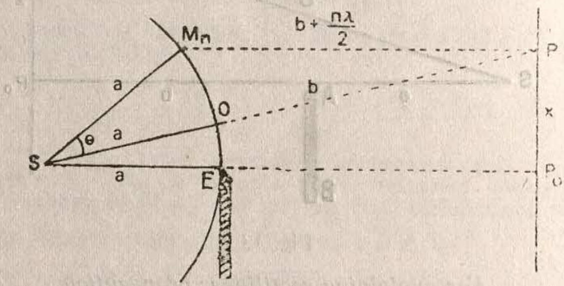


Fig. 13·8.

For diffraction at a straight edge

the plane of paper. As a result, $M_1, M_2 \ldots M_n$ draw lines on the cylindrical wavefront. The areas between these lines define the *h.p.* zones. These are strips, not rings, and Fig. 13·9 represents them on both sides of O. Notice that the waves from $1', 2', 3', 4', \ldots$ reach P in the same phase as those from $1, 2, 3, 4 \ldots$ respectively. In that sense, strips n and n' taken together form the n^{th} *h.p.* zone.

The width w_n of n^{th} *h-p.* zone may be calculated from eq. (13·6) which gives :

$$w_n = r_n - r_{n-1} = \sqrt{\frac{ab\,\lambda}{a+b}}\left[\sqrt{n} - \sqrt{n-1}\,\right] \qquad \ldots(13{\cdot}14)$$

The area of the *h.p.* zone is proportional to w_n and it decreases as n increases. (For $n \geqslant 1$, w_n is propartional to $n^{-1/2}$).

The strip zones shown in Fig. 13.9 when placed with Fig. 13·8 will have their lengths along perpendicular to the plane of paper. Then *within a given h.p. zone strip* the different parts will have different distances from the observation point P. To deduce the resultant disturbance at P due to any one strip, like strip 5 in Fig. 13·9, we may sub-divide the strip into sub-half-period zones. The amplitude of the resultant disturbance at P due to the whole *h.p.* zone strip is then given by some *fraction* of the amplitude due to the central sub-zone (shown shaded). This value of the amplitude of disturbance at P due to the n^{th} *h.p.* strip (taken collectively) we represent by R_n. Assuming the fraction to be the same for all *h.p.* zone strips, we may put, in view of eq. (14·14),

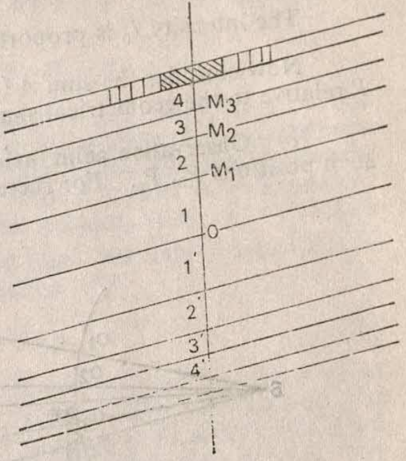


Fig. 13·9

The h.p. strip zones on a cylindrical wave-front

$$R_n \propto (\sqrt{n}-\sqrt{n-1}) \qquad \ldots(13{\cdot}15)$$

The disturbances due to sucessive half-period strips arrive at P with successive phase differences π. Hence the resultant amplitude at P is given by

$$A=(R_1-R_2+R_3\cdots)+(R_1-R_2+R_3\cdots\cdots) \qquad \ldots(13{\cdot}16)$$

where we put two series because there are *h.p.* strips on both sides of the pole O. As many zones as are open should be included in evaluating A.

n	R_n	S_n
1	1·000	1·000
2	0·414	0·586
3	0·318	0·904
4	0·268	0·636
5	0·236	0·872
6	0·218	0·659
7	0·196	0·855
8	0·183	0·672
...	....	....
...	....	....
∞	0·000	0·77

It is necessary to remember that (unlike the case of spherical wavefront) the magnitudes of $R_1, R_2\cdots\cdots$, decrease rapidly as seen from eq. (13·14). Table below gives relative values of the R's for different values of n, and the sum (S_n) of one series in eq. (13·16) upto n terms, taking $R_1=$ 1·000 unit.

When there is no obstacle, *h.p.* zones upto very large n are effective. Each sum in eq. (13·17) then tends to $0{\cdot}77R_1$. Hence, calling the amplitude at P due to the whole wavefront as A_0 we get

$$A_0 = 0.77R_1 + 0{\cdot}77R_1 \approx 1{\cdot}54\ R_1. \qquad \ldots(13{\cdot}17)$$

The intensity I_0 is proportional to A_0^2.

Now consider the sum A (eq. 13·16) for different positions of P relative to the geometrical shadow P_0 of the edge E.

(*b*) **Observation point inside the shadow.** Fig. 13·10 shows two such positions P_3, P_4. For these position all the lower *h.p.* strips

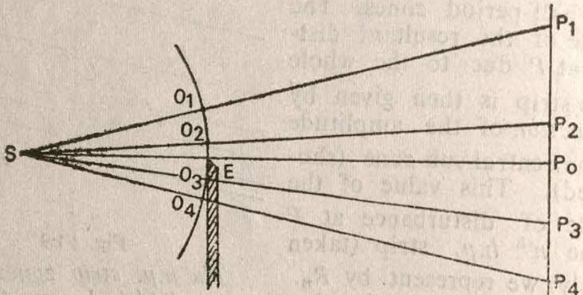


Fig. 13·10.

For considering intensity at various points relative to P_0

are cut off, as also a number (say n) out of the upper *h.p.* strips. So we get from eq. (13·16)

$$|\,A\,| = (R_{n+1} - R_{n+2} + R_{n+3} - \ldots\ldots)$$
$$= \text{some fraction of } R_{n+1}$$

When P is deep inside the shadow, n is large, hence R_{n+1} is very small (see Table above). Thus point P is dark. Now, as P approaches the edge of the geometrical shadow, n decreases. Hence R_{n+1} *increases continuously* and so does the intensity at P.

(*c*) **At the edge of the shadow.** In this case the lower *h.p.* strips are all cut, upper strips are all allowed.

$$\therefore \quad A = (R_1 - R_2 + R_3 \ldots\ldots) \sim 0{\cdot}77R_1$$

Comparing with eq. (13·17)

$$\frac{I}{I_0} = \frac{A^2}{A_0^2} = \frac{1}{4} \qquad \ldots(13{\cdot}18)$$

Thus the intensity at the edge P_0 is neither zero nor the full value I_0. It is one quarter of I_0.

(*d*) **P outside the shadow.** Fig. 13·10 shows two such positions P_1, P_2. Now, all upper *h.p.* zones or strips are allowed, plus some zones (say n) from the lower ones also. Then the amplitude at P is given by

$$A=(R_1-R_2+R_3\ldots\ldots\text{to }\infty)+(R_1-R_2+R_3\ldots\ldots\text{to }R_n)$$

The sum in the first series is fixed at 0·77 R_1, while that in the second series becomes large or small according as n is odd or even respectively. Thus as P moves away from P_0 the intensity alternately rises and falls. Also the variations become smaller as n increases—till it reaches the steady value I_0 (corresponding to $A=A_0=1{\cdot}54\ R_1$).

The number (n) of half-period zones corresponding to $PP_0=x_n$ is given by eq. (13·13), which leads to

$$\left[\frac{\lambda b(a+b)}{a}\right]^{1/2}\sqrt{n}=K\sqrt{n} \qquad (13{\cdot}19)$$

where K is constant for a given set-up. The positions of the minima and maxima are given by n even and n odd respectively.

We can now draw the intensity distribution curve. Fig. 13·11 (top) shows the graph of intensity I against distance x measured from P_0. It shows that

(*i*) Inside the shadow there are no fringes, but gradual fall of intensity.

(*ii*) At $x=0$, intensity$=\frac{1}{4}I_0$.

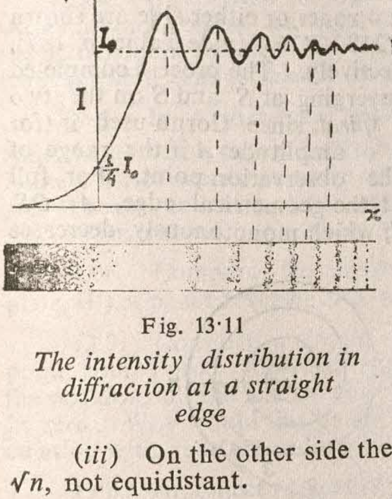


Fig. 13·11

The intensity distribution in diffraction at a straight edge

(*iii*) On the other side there are fringes, spaced according to $\sqrt{n}$, not equidistant.

(*iv*) The fluctuations of intensity become weaker as x increases. At large x the intensity reaches a steady value I_0.

Example 13·7. In an experiment with straight edge diffraction, the slit to edge distance is 1·00 metre, and edge to screen distance is 2·00 metres. If $\lambda=5000A$, deduce the separation of the first three bright fringes.

Solution. The *positions* of the maxima (relative to geometrical projection of the edge) are given by eq. (13·19)

$$x_m=K\sqrt{n}, \qquad \text{where } n=1,3,5,\ldots\ldots$$

and $$K=\left[\frac{b(a+b)\,\lambda}{a}\right]^{\frac{1}{2}}$$

We have $a=100$ cm, $b=200$ cm, $\lambda=5\times10^{-5}$ cm.

$$\therefore \qquad K=\left(\frac{200\times300\times5\times10^{-5}}{100}\right)^{\frac{1}{2}}=0{\cdot}171\text{ cm.}$$

The first three maxima therefore lie at

$$0{\cdot}171, \qquad 0{\cdot}171\sqrt{3}, \text{ and } 0{\cdot}171\sqrt{5}\text{ cm.}$$

The separations are

$$0{\cdot}171\sqrt{3}-0{\cdot}171=0{\cdot}125 \text{ cm.}$$

$$0{\cdot}171\sqrt{5}-0{\cdot}171\sqrt{3}=0{\cdot}086 \text{ cm.}$$ ⊙

13·7. Cornu's Spiral. Instead of $\frac{1}{2}$-period zones, one could draw $\frac{1}{2m}$-period zones on the cylindrical wavefront, where m may be 10 or 100 or ∞ in the extreme case. The relative areas of the successive zones for a cylindrical wavefront would still be proportional to $\sqrt{n}-\sqrt{n-1}$; but their contributions would be added with phase differences π/m (not π). Algebraic addition like eq. (13·1) would then be inapplicable.

Fig. 13·12 shows the *phasor diagram* addition for the case $m=8$. The contributions from first 4 zones or either side are shown in the amplitude ratio 1 : 0·414 : 0·318 : 0·268 (vide Table, p. 187), with phases 0, $\pi/8$, $2\pi/8$, $3\pi/8$ respectively. The process completed for all zones leads to the spiral converging at S' and S on the two sides. The spiral is called *Cornu's spiral*, since Cornu used it (for $m \to \infty$) for quantitative evaluation of amplitude A if the range of open wavefront was known for the observation point. For full wavefront open, $A=S'S$. For P at the geometrical edge, $A=OS$. As P moves in the shadow, $A=XS$, which monotonously decreases as X goes up the spiral. As P moves in the open side, $A=X'S$, which alternately becomes large and small as X' goes down the spiral. (cf. Fig. 13·11).

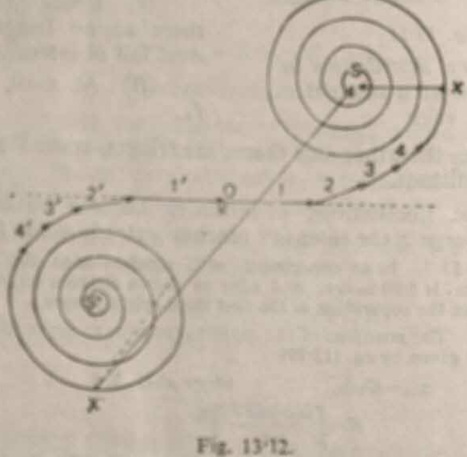


Fig. 13·12.
Cornu's Spiral

PROBLEMS

13·1 Describe two simple observations which you can make in everyday life to show that light is diffracted.

13·2. Consider a wavefront divided into one-fourth period zones. If amplitude contributed at the observation point is independent of the zone number *n*, find by the phasor method the resultant amplitude due to the first (*i*) 2 zones, (*ii*) 4 zones, (*iii*) 6 zones, (*iv*) 8 zones. Repeat taking amplitude varying as $n^{-1/2}$.

13·3. Deduce the area of a h.p. zone for $a=30$ cm, $b=150$ cm, $\lambda=6\times10^{-5}$ cm in case of spherical wavefronts. Deduce the two smallest radii of a circular aperture for which the observation point will be (*i*) dark, (*ii*) bright.

13·4. If all parts of a given h.p. zone are assumed to send waves in same phase, the amplitude contribution is *R*. Show that the actual contribution will be $2R/\pi$ if the phase variation for waves from within a zone is taken into account.

13·5. In a cylindrical wavefront the successive equal-phase zones have areas proportional to $\sqrt{m+1}-\sqrt{m}$, where $m=0, 1, 2,\ldots$ Make a list of successive areas upto $m=8$. Now consider *one-eighth* period zones, and deduce the resultant relative intensity due to 1, 2, 3,......8 such zones. (Compare with Problem 13·2).

13·6. Compare the image forming capacity of (*i*) a zone plate, (*ii*) a phase reversal zone plate, and (*iii*) a lens.

13·7. Consider a circular aperture of radius 1·0 cm. For a point source of $\lambda=5000A$ at infinite distance along the axis, deduce the positions on the axis on the other side where the intensity would be zero. Why would this be difficult to observe ? Repeat the calculation with aperture radius $=0·1$ cm.

13·8. An object placed 50 cm away from a zone plate gives a real image at 70 cm on the other side for $\lambda=5000A$. Calculate (*i*) position of the image for $\lambda=7000A$, (*ii*) positions of the next two secondary images for $\lambda=5000A$, (*iii*) approximate intensities of secondary images relative to the primary.

13·9. Deduce the principal focal length of a zone plate whose central zone has radius 0·62 mm and light used has $\lambda=5000A$. Deduce also the outer radius of 9th h.p. zone.

13·10. A point source of $\lambda=6000A$ is placed 50·0 cm away along the axis of a circular aperture 4·0 mm in dia. Locate the axial position on the other side so that the aperture covers 16 *h.p.* zones. Calculate also the diameters of the first and second bright rings around the central spot there.

13·11. Show that *departures* from the rectilinear propagation are limited to a *small* region near the edges of the geometrical shadow. Discuss the extent of *smallness* of this region, and the nature of *departures*.

13·12. A narrow slit is illuminated with light of $\lambda=5{\cdot}89\times10^{-5}$ cm. A straight edge is placed 2·0 m away, the edge being parallel to the slit. Observations are made 3·0 m from the edge. Compute the positions of the first three maxima and the intervening minima relative to the 'geometrical edge' of the shadow.

13.13. Figure below is a spiral and it gives the phasor diagram for a spherical wavefront. Explain, why this differs so much with the phaser diagram for a cylindrical wavefront (Cornu's spiral).

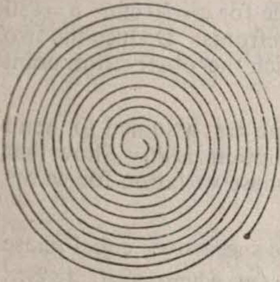


Fig. 13·13.

13·14. The Figure in Q. 13·13 is not quantitative. Yet, treat it as quantitative and from geometrical measurements deduce (in terms of I_0 for the full wavefront) the intersity I for (*i*) central 7 h.p. zones open, (*ii*) central 8 h.p. zones open, (*iii*) central 10 h.p. zones blocked, rest open.

13·15. From the phasor in Q. 13·13, how would you pass on to the phasors of (*i*) a zone plate, (*ii*) a phase reversal zone plate, (*iii*) an ideal lens ?

14

Diffraction of Light (II)

14·1. Fresnel and Fraunhofer Classes of Diffraction. In Fig. 14·1 (top), $A'B'$ and $C'D'$ are *projections of the apertures AB* and *CD* which limit the wavefront. The gross structure of $A'B'$ (for instance) matches AB but has nothing to do with the structure of the source S. The modifications in detailed structure of $A'B'$ are called *Fresnel diffraction*, which we studied in Chapter 13.

In Fig. 14·1. (bottom), lenses L_1 and L_2 are added so that S' is an *image of the source S* Now, the gross struture of S' has nothing to do with the structure of AB or CD. The modifications in the detailed structure of S' due to apertures which limit the wavefront are called *Fraunhofer diffraction*, which we will now study.

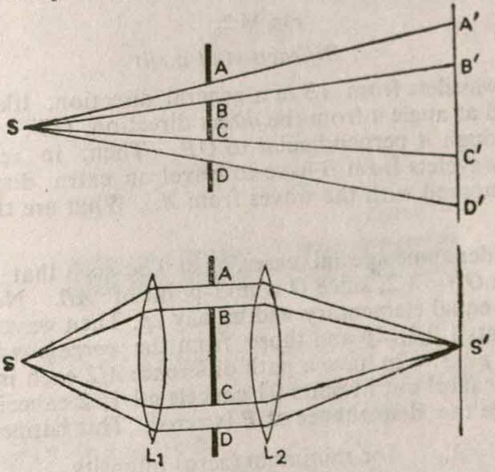


Fig. 14·1.

Fresnel and Fraunhofer diffractions

It is of interest to note that if in Fig. 14·1 (lower) lens L_1 or L_2 or both are removed and correspondingly S or S' or both are shifted to infinite distance, the situation in effect does not change. For this reason, Fraunhofer diffraction case is also defined as that in which the source as well as the plane of observation are at infinite distance from the diffracting element. As one result, the *angular inclinations* are important here, whereas lateral *distances* are important in the Fresnel class.

As we shall see, Fraunhofer class of diffraction can be treated in rigorous mathematical terms and has wide-ranging applications.

14·2. Fraunhofer Diffraction at a Single Slit : (*a*) Qualitative Treatment. In Fig 14·2 AB is a slit of width a, and its length is perpendicular to the plane of paper and is large (treated as in finite). Source point S and lens L_1 make plane wavefronts which reach the slit, and lens L_2 produces an image S', which should be a point image of S according to geometrical optics. To discuss how, according to the wave theory, the finite width a modifies the image, consider secondary wavelets from AB in a general direction, like a direction OP inclined at angle θ from the *direct* direction, OS'. Let ANM be a plane through A perpendicular to OP. Then, in reaching P the secondary wavelets from B have to travel an extra distance $BM=i$ sin θ as compared with the waves from A. What are the consequences ?

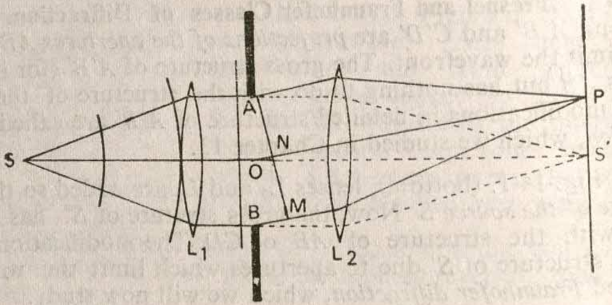


Fig. 14·2.

Diffraction at a slit

Consider some special cases. Let θ be such that $BM=a \sin\theta=\lambda$. Then $ON=\lambda/2$, since O is mid-point of AB. Now divide AB into several equal elementary widths, say $2p$. Then waves from successive elements $1,2,3,\cdots p$ and those from the *corresponding* elements $p+1, p+2, p+3\cdots 2p$ have a path difference $\lambda/2$ each in reaching P. Hence they cancel out in pairs : 1 cancels $p+1$, 2 cancels $p+2\cdots$ and so on. Hence the disturbance at P is zero. This happens when

$$a \sin\theta=\lambda, \quad \text{for minimum (zero) intensity} \qquad \ldots(14{\cdot}1)$$

Thus, *when the path difference at P between extreme wavelets from a given opening becomes λ, the intensity at P becomes zero.*

This is not to be confused with interference between waves from two coherent slit sources A and B. There the intensity is zero when path difference is $\lambda/2$. Here we have one continuous slit of width A to B.

If θ is such that $a \sin \theta = 3\lambda/2$, we can divide the slit width AB into three parts. The waves from successive elements from the first part and those from the *corresponding* elements from the second part mutually cancel out. We are thus left with waves from one-third of the slit only. There are path differences from 0 to $\lambda/2$ between the extremes within this one part too ; if we omit detail and consider all these wavelets in one phase, then the amplitude corresponding to this angle is $\sim 1/3$ times of that at S' ; the intensity is $\sim 1/9$ times. We put the condition

$$a \sin \theta = \frac{3\lambda}{2} \text{ for next maximum (intensity } \sim \frac{I_0}{9}). \qquad \ldots(14{\cdot}2)$$

We generalise results (14·1) and (14·2) to say that

$$a \sin \theta = \lambda, 2\lambda, 3\lambda \ldots \text{minima (zero intensity)} \qquad \ldots(14{\cdot}3a)$$

$$a \sin \theta = \frac{3\lambda}{2}, \frac{5\lambda}{2} \ldots \text{maxima (intensity} \sim \frac{I_0}{9}, \frac{I_0}{25} \ldots) \qquad \ldots(14{\cdot}3b)$$

Notice that Eq. 14·3*a* and 14·3*b* are equivalent to dividing the open wavefront into *even* and *odd* number of half-period zones, respectively. The difference front Fresnel's case is that angles are involved here, and rigorous deductions are possible (as you will see).

In Fig. 14·3 we show the intensity distribution curve, as well as the schematic diffraction pattern observed.

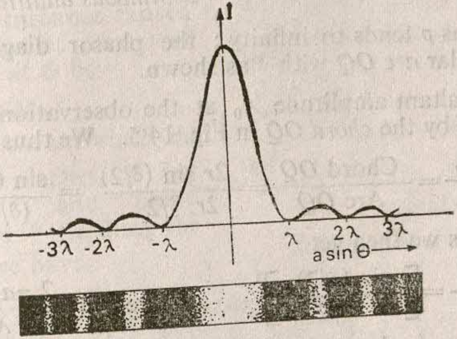


Fig. 14·3

Intensity pattern in a single slit diffraction

We draw particular attention to the fact that $a \sin \theta = n\lambda$ with $n = 0$ does not give a minimum but a maximum. This is the centre of the image (S' in Fig. 14·2).

(*b*) **Phasor Diagram Method.** In Fig. 14·2 the addition of the secondary wavelets from different parts of slit AB as they reach P can be done by the phasor diagram method (§ 11·2).

Let us divide slit AB into p *equal* parts. Then the amplitude contributed by each element is A_0/p, if A_0 is that if the entire slit sends waves in the same phase (as at S'). Also, the phase difference between waves for successives elements is δ/p, where

$$\delta=\frac{2\pi}{\lambda}(a \sin \theta) \qquad \text{...(14·4)}$$

Fig. 14.4 now shows the phasor diagram to add p disturbances of amplitude A_0/p each and succassive phase disturbanees δ/p. Fig. 14·5

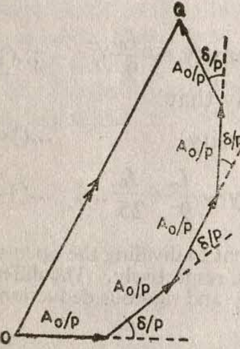


Fig. 14·4.

Stepped addition of wavelets

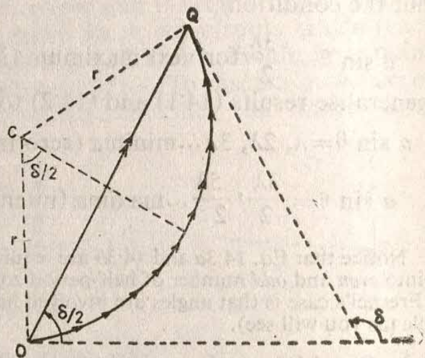


Fig. 14·5.

Continuous addition of wavelets

is the result as p tends to infinity ; the phasor diagram becomes a smooth circular *arc OQ* with δ as shown.

The resultant amplitude A_0 at the observation point P (Fig. 14·2) is given by the *chord OQ* in Fig. 14·5. We thus have

$$\frac{A_\theta}{A_0}=\frac{\text{Chord } OQ}{\text{Arc } OQ}=\frac{2r \sin (\delta/2)}{2r.\,\delta/2}=\frac{\sin (\delta/2)}{(\delta/2)} \qquad \text{...(14·5)}$$

For intensities we then get

$$\frac{I\theta}{I_0}=\left[\frac{\sin (\delta/2)}{\delta/2}\right]^2 \qquad \text{where} \quad \delta=\frac{2\,\pi a \sin \theta}{\lambda} \qquad \text{...(14·6)}$$

$$=\left(\frac{\sin p}{p}\right)^2 \qquad \text{where} \quad p=\frac{\pi a \sin \theta}{\lambda} \qquad \text{...(14·6}a\text{)}$$

How the phasor diagram changes with increasing θ (hence increasing δ) is shown in Fig. 14·6. For $\delta=0$ (observation point S' in Fig. 14·2) the phasor curve is a straght line of length A_0. For $\delta=\pi$ it becomes a semicircle, for $\delta=2\pi$ a circle, for $\delta=3\pi$ it goes round one circle and a half, for $\delta=4\pi$ two full circles. Thus $\delta=2\pi, 4\pi, 6\pi$......correspond to chord $OQ=0$, which means zero amplitude (hence zero intensity). Also, $\delta=3\pi$, 5π......correspond to (nearly) secondary maxima, the main maximum being at $\delta=0$.

* Notice that sin p/p is the sinc function referred to in § 9·3. Discussion of the relation will lead us outside the scope of this book.

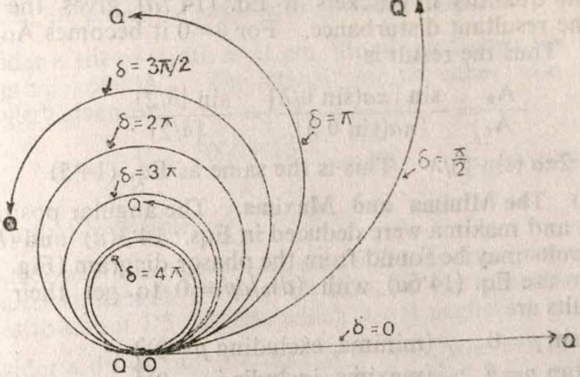


Fig. 14.6.

Showing how the phasor diagram curls round as δ increases

(c) **Method of Integral Calculus.** The addition of the secondary wavelets reaching P from different parts of slit AB (Fig. 4·2) can also be done by integration. Fig. 14·7 may be used for this.

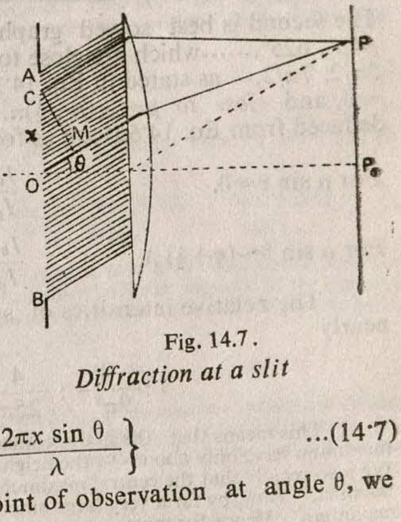


Fig. 14.7.

Diffraction at a slit

Let the disturbance caused at P by the wavelet *from unit width* of the slit at O be

$$y_0 = A\cos\omega t$$

Then the wavelet from width dx at C when it reaches P has amplitude Adx and phase $\omega t + (2\pi/\lambda)x\sin\theta$. Calling the disturbance dy we have

$$dy = Adx\cos\left\{\omega t + \frac{2\pi x\sin\theta}{\lambda}\right\} \quad \ldots(14{\cdot}7)$$

For the total disturbance at the point of observation at angle θ, we get

$$y = \int_{-a/2}^{+a/2} A\cos\left\{\omega t + \frac{2\pi x\sin\theta}{\lambda}\right\}.dx$$

$$= \left[A.\frac{\sin(\pi a\sin\theta/\lambda)}{\pi\sin\theta/\lambda}\right]\cos\omega t^{*} \quad \ldots(14{\cdot}7a)$$

---

* $y = A\cos\omega t\int_{-a/2}^{+a/2}\cos\frac{2\pi x\sin\delta}{\lambda}dx - A\sin\omega t\int_{-a/2}^{+a/2}\sin\frac{2\pi x\sin\delta}{\lambda}dx.$

The second integral is zero. The first gives the result (14·7 a)

The quantity in brackets in Eq. (14·7a) gives the amplitude A_θ of the resultant disturbance. For $\theta=0$ it becomes Aa, which we call A_0. Thus the result is

$$\frac{A_\theta}{A_0}=\frac{\sin\left[\pi a(\sin\theta/\lambda\right]}{\left[\pi a(\sin\theta/\lambda\right]}=\frac{\sin\left[\delta/2\right]}{\left[\delta/2\right]}$$

with $\delta=2\pi a\,(\sin\theta)/\lambda$- This is the same as Eq. (14·5).

(*d*) **The Minima and Maxima.** The angular positions of the minima and maxima were deduced in Eqs. 14·3(a) and (b), and the same results may be found from the phasor diagram (Fig. 14·6). One may also use Eq. (14·6a) with $(dI_\theta/dp)=0$ to get their positions. The results are

$$\left.\begin{array}{ll}\sin p=0 & \text{(minima, excluding } p=0)\\ \tan p=p & \text{(maxima, including } p=0)\end{array}\right\} \qquad (\ldots 14{\cdot}8)$$

The first of these gives Eq. (14·3a), which we repeat :

$$\text{minima at } \boxed{a\sin\theta=n\lambda}\,,\ n=\pm1,\ \pm2\cdots \qquad \ldots(14{\cdot}9)$$

The second is best solved graphically*, and it leads to $p=258°$, 443°, 625°,......which are close to, though not exactly equal to $3\pi/2$, $5\pi/2$, $7\pi/2$,... as stated in Eq. 14·3 (b). Thus the maxima fall at $p=0$, and *close to* $p=\pm(n+\frac{1}{2})\pi$. The intensities at the maxima, deduced from Eq. 14·6a) are as follows :

For $a\sin\theta=0$. $\qquad \dfrac{I_\theta}{I_0}=\left(\dfrac{\sin p}{p}\right)^2=1$

For $a\sin\theta\sim(n+\frac{1}{2})\,\lambda$. $\qquad \dfrac{I_\theta}{I_0}=\left(\dfrac{\sin p}{p}\right)^2\sim\dfrac{1}{[\pi(n+\frac{1}{2})]^2}$

The relative intensities of successive maxima are, therefore, nearly

$$1:\frac{4}{9\pi^2}:\frac{4}{25\pi^2}:\frac{4}{49\pi^2}\ldots \qquad \ldots(14{\cdot}10)$$

This means that the first secondary maxima on either side of the central maximum have only about 4% the height of the central maximum (See Fig. 14·3). We also recall that the central maximum has double the width of the secondary maxima. Thus, by far a very large amount of light falls in the central diffraction maximum. Hence for most purposes we may just discuss the *spread* of the central maximum and neglect the secondary maxima.

(*e*) **Width of the Central Maximum.** The most significant result of diffraction is the finite angular width of the central maximum given by Eq. (14·9) with $n\pm1$. It means that the energy passing through the slit spreads out upto angle θ given by

$$\theta=\pm\sin^{-1}\frac{\lambda}{a} \qquad \ldots(14{\cdot}11)$$

---

[*] Draw curves for equation $f(p)=\tan p$, and for $f(p)=p$. The intersection points give values of p which satisfy the equation $\tan p=p$.

The theory is applicable for all kinds of waves—sound waves, waves over liquid surface, and electromagnetic waves ranging from radio waves to gamma rays. Of course, the effects are meaningful only when a is not$\gg \lambda$.

Consider a slit of width $a=1$ cm, illuminated normally by light of wavelength $\lambda=5000\text{A}^\circ$. The intensity on the other side spreads upto an angle θ given by

$$|\theta|=\frac{\lambda}{a}=\frac{5000\times 10^{-8}}{1}\text{ radian}\approx\frac{1}{300}\text{degree}$$

which is hardly observable.

Consider now the case $a={\cdot}02$ mm. (This is the order of width used in a spectrometer slit). For this case, for the same wavelength, $|\theta|$ comes to about 1·5 degrees, which is not negligible.

Consider a diffraction grating (see § 14·6), where each opening has width $\sim 6\times 10^{-5}$ cm. The diffraction 'spread' θ on either side of the normal is now given (for $\lambda=5000\ \text{A}^\circ$) by $\sin\theta=\pm 0{\cdot}83$, which gives $|\theta|=56^\circ$. Thus the intensity in this case spreads upto quite a large angle

Example 14·1. Sound waves of frequency 3000 sec^{-1} fall normally on an opening of width ·50 metre. Calculate the angular positions of the diffraction minima. [Take speed of sound as 345 m/sec.]

Solution. $\lambda=\frac{345}{3000}=0{\cdot}115$ m, $\frac{\lambda}{a}=0.23$

Minima occur at $\theta=\pm\sin^{-1} n\lambda/a$. Since maximum value of $\sin\theta$ can be the only minima are at

$$\pm\sin^{-1}0{\cdot}23,\quad \pm\sin^{-1}0{\cdot}46,\quad \pm\sin^{-1}0.69,\quad \pm\sin^{-1}0{\cdot}92$$

or

$$\pm 13^\circ,\quad \pm 27^\circ,\quad \pm 44^\circ,\quad \pm 67^\circ$$

[Note. Note if an observer walks across at a large distance D from the opening, then starting from a point on the perpendicular bisector of the opening he will observe minima of sound at *positions* given by $D\tan\theta$. If for example, $D=4$ metre, the mimina appear at

$$\pm 0{\cdot}92\text{ m},\quad \pm 2{\cdot}0\text{ m},\quad \pm 3.8\text{ m},\quad \pm 9{\cdot}9\text{ m}]$$

Example 14·2. Microwaves of 6000 Mc/s frequency are incident normally on a slit of variable width and large length. A receiver is placed at a large distance on the other side at angle 20° with the normal. For what widths of the slit would the receiver show zero intensity ?

Solution. Speed of microwaves=speed of light $=3\times 10^{10}$ cm/sec.

Hence

$$\lambda=\frac{c}{\nu}=\frac{3\times 10^{10}}{6\times 10^{9}}=5{\cdot}0\text{ cm}$$

Let the variable slit width be a. Then for minima

$$a\sin 20^\circ=n\times 5{\cdot}0\,;\qquad (n=\pm 1, \pm 2\ldots)$$

$\sin 20^\circ=0{\cdot}34$. Substitution gives

$$a=\frac{5n}{0{\cdot}34}=15,\ 29,\ 44\ldots\ldots\text{cm}$$

14·3 Fraunhofer Diffraction as Fourier Transform of the Aperture. The similarity between the Fourier transform of a square pulse, Eq. (9·33), and the Fraunhofer diffraction at a single slit, Eq. (14·5), may be noted. This is not accidental. A 'slit' may be described by a transmittance function whose value is unity over a width a and zero elsewhere. Fig. 9·4 (left) with $h=1$ and $b=a$ is therefore a pure mathematical description of a 'slit'. Then, if the aperture a is infinite, its fourier transform, Fig. 9·4 (right), has zero spread ; correspondingly, the Fraunhofer diffraction has *zero* spread. As a narrows down, the diffraction spread increases.

A further consideration is of interest. Let a rectangular slit of length a along x-axis and b along y-axis intercept a plane wavefront (from a point source) travelling along the z-axis. Its Fraunhofer diffraction will show spreads along x- and y-directions in inverse proportions to a and b respectively. Fig. 14·8 gives a schematic representation. (The dimensions λ/a and λ/b are in angles ; in a photographic record they will have linear dimensions f times, where f is the focal length of the system). It is both exciting and educative to view a distant halogen lamp through apertures with varions a, b combinations.

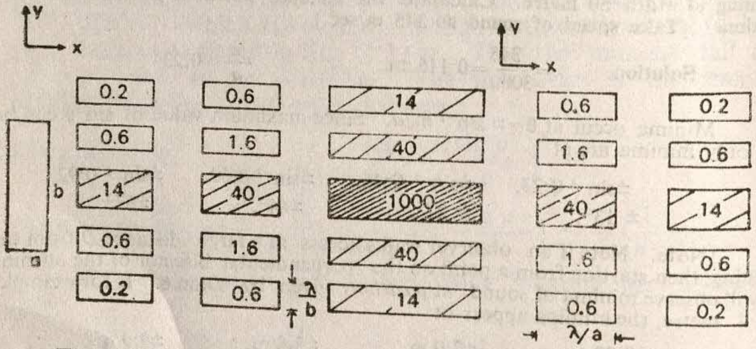


Fig. 14·8. *A slit with bo th a and be finite, and its Fraunhofer diffraction (Schematic) (leftend) (right). The Fraunhoper diffraction pattern (Schematic). Figures in the rectangles show very approximate relative peak intensities in the different maximum, using Eq. (14·10).*

14·4 Fraunhofer Diffraction at a Circular Aperture. We have discussed the diffraction at a narrow slit (a finite and b infinite), and a rectangular aperture (a and b both finite). Figs. 14·3 (bottom) and 14·8 show the corresponding Fraunhofer patterns. The case of a circular aperture is of symmetry about the z-axis, which means indistinguishability of x and y axes. A rigorous treatment of this case is difficult (one needs a series solution) and we will not give it here. But there are three ways of visualising the sitution :

(*a*(Rotate Fig 14·3 (bottom) about an axis passing through the centre and perpendicular to the plane of paper. You get alternate bright and dark rings. (Why rotate ? The answer is not straight.)

(*b*) If *AB* in Fig, 14·7 is not a 'slit', but a circular aperture of *diameter a* then the strip of width dx taken at C has a finite length $(a^2-4x^2)^{1/2}$, which varies form a at $x=0$ to 0 at $x=\pm a/2$. This factor must be included in the expression (14·7) for dy, The step is rigorous, but integrating over $x=-a/2$ to $-a/2$ is not simple.

(*c*) In Fig. 14·8, consider $a=b$ case, i.e., a square aperture. Then consider how the diffraction pattern may change if the corners of the apertures are blocked-off to make it nearer a circular one.

The result given by rigorous theory is that the central images is a disc of angular radius θ, given by

$$a \sin \theta_1 = 1.22\lambda \tag{14·12}$$

It means that intensity falls off with increasing θ, reaching a minimum at $\theta=\theta_1$. Thereafter it rises, reaches a certain maximum (much less than that at $\theta=0$), falls again and so on. We are not interested in these secondary maxima and minima at present, particularly because they contain a very small fraction of the incident energy.

The essence of wave theory is that the image of any distant point source (like a star) made by the most ideal lens of focal length *f* will not be a point image, but a disc image whose dimensions are given by

$$\text{angular radius } \frac{1{\cdot}22\lambda}{a}; \quad \text{actual radius } \frac{1{\cdot}22\lambda}{a} f.$$

This result is of far reaching significance in image forming instruments, and we shall use it in Chapter 15 in some detail.

14.5. Emission by a Disc Source. The result of Eq. (14·12) is of much wider significance than just transmission. It applies to emission as well as reception of waves. Consider a source of *sound* waves comprising a disc of diameter a and oscillating as a whole normal to its face. Then the maximum intensity will be along the normal (as expected), but the sound will necessarily spread upto a cone of semi-angle $\sin^{-1}(\lambda/a)$ nearly. It $\lambda > a$, the spread will extend upto 90° ; on the other hand if $\lambda = a/20$ (say), the spread will be limited to ~ 1/20 radian about the normal. We can reverse the statement. In order to limit the spread of the emitted waves to a cone of semi-angle θ (considered small), the diameter over which we must have in-phase (coherent) oscillations should be $\sim\lambda/\theta$. The statement applies to radar antenna and receivers also (see § 15·7).

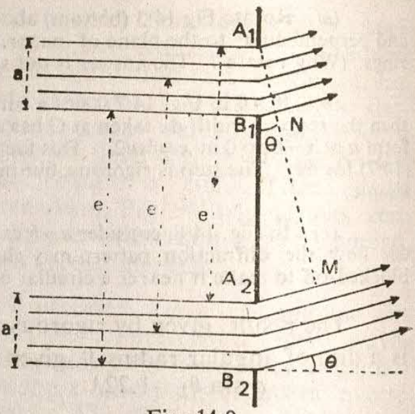


Fig. 14·9.
Two-slit diffraction

14·6. Diffraction at Two Parallel Slits. In Fig. 14·9, let A_1B_1 and A_2B_2 be two slits of width a each and separation e*. Lengths of the slits are perpendicular to the plane of paper. Consider monochromatic light of wavelength λ falling normally on the slits from the left. All the wavelets proceeding in a *direction* θ from the normal are collected by a lens at a given point in its focal plane. What is the resultant intensity and how does this vary with θ ?

We here have *interference* of two waves of amplitude A_θ each, having a phase difference $\Delta = (2\pi/\lambda)\, e \sin\theta$. The resultant amplitude R_θ is therefore given by

$$R_\theta^2 = A_\theta^2 + A_\theta^2 + 2A_\theta A_\theta \cos\Delta = 2A_\theta^2 (1+\cos\Delta)$$
$$= 4A_\theta^2 \cos(\Delta/2).$$

A_θ is the result of *diffraction* at each slit (width a), and it is given by Eq. (14·5). Substituting for A_θ we get, for intensity.

$$I_\theta = 4I_0 \left(\cos\frac{\Delta}{2}\right)^2 \left(\frac{\sin\delta/2}{\delta/2}\right)^2 \qquad \text{...(14·13)}$$

where $\Delta = (2\pi/\lambda)\, e \sin\theta$, and $\delta = (2\pi/\lambda)\, a \sin\theta$. ...(14·13a)
and I_0 = intensity due to a single slit in the direction $\theta = 0$.

Let us consider as example a case where $e = 4a$, which makes $\Delta = 4\delta$ for all angles (see Eq. 14·13a). In Fig. 14·10 the upper curve is plot of the diffraction term against δ and is the same as in Fig. 14·3 except that phase is the x- axis here. (The *minima* occur at $\delta = 2\pi, 4\pi$... on each side of the normal). The middle curve is a plot of the interference term against Δ. This is a $\cos^2 \Delta/2$ curve with *maxima* at $\Delta = 0, 2\pi, 4\pi$,...on each side. The lower curve is a plot of the product of the ordinates in the upper two curves, *i.e.*, of I_θ/I_θ.

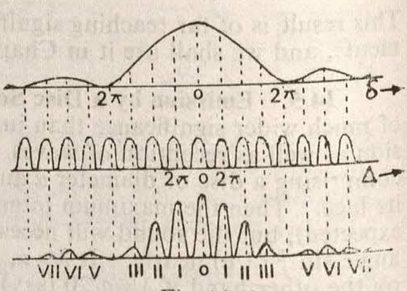


Fig. 14·10.
Intensity distribution due to two parallel slits. Top curve : the diffraction term ; middle curve : the interference term ; bottom curve : the resultant distribution

We note in particular that

---

† e is the distance of *corresponding* parts in A_1B_1 and A_2B_2. If opaqe par
$B_1A_2 = b$, then $e = a + b$.

the central diffraction maximum is divided now into several interference fringes, whose maxima ocur at

$$\triangle = n.\,2\pi \qquad i.e., \qquad e \sin\theta = n\lambda, \qquad ...(14{\cdot}14)$$

where $n=0, 1, 2...$ The value of n defines the "*order*" *of the interference maximum*, marked 0 and I, II, III...on both sides in Fig. 14·10. In the case taken by us ($e=4a$), the first diffraction maximum covers the *0 order* and 3 more *orders* on each side. The fourth order falls at the zero intensity of diffraction. It is therefore absent.*

Example 14·3. Two parallel slits have widths 10 mm each and opaqeu space between them is 20 mm. If light of wavelength 5 mm (this is in microwave region) falls normally on the system, deduce (i) the angular spread of the central diffraction maximum, (ii) the angular positions of the first 4 interference maxima on one side.

Solution. $a=1{\cdot}0$ cm. $e=3{\cdot}0$ cm. $\lambda=0{\cdot}5$ cm.

Spread of central diffraction maximum is given by

$$\theta = \pm\sin^{-1}(\lambda/a) = \pm 30^\circ.$$

The interference maxima are given by Eq. (14·14) as

$$\theta = \sin^{-1}(n\lambda/e) = \sin^{-1}(n/6).$$

Apart from the zero order maximum ($\theta=0$), the four maxima on one side occur at

$$\theta = \sin^{-1}(1/6),\ \sin^{-1}(2/6),\ \sin^{-1}(3/6),\ \sin^{-1}(4/6)$$

$$= 9^\circ 36', \qquad 19^\circ 30', \qquad 30^\circ 0, \qquad 41^\circ 48'.$$

However, the interference 'maximum' due to appear at 30° has zero intensity since the first 'zero' of diffraction falls there. ⊙

14.7. Diffraction at N Parallel Slits. (*a*) **The intensity distribution.** In Fig. 14·11(*a*) a number (*N*) of equidistant and equally

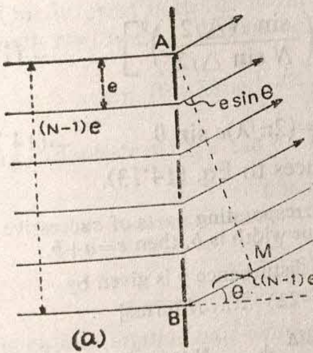


Fig. 14·11 (a)
N parallel slits

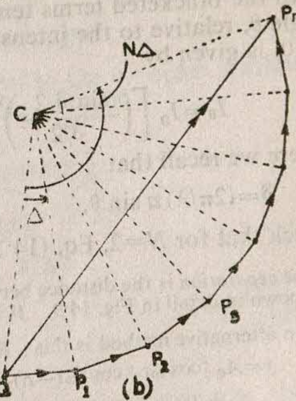


Fig. 14·11 (b)
Summing N waves of equal amplitudes and equal phase differences

---

* In general, if $e=(p/q)\,a$, with p and q as integers, then q^{th} order diffraction minimum ($a \sin\theta = q\lambda$) coincides with p^{th} order interference maximum (e sin $\theta = p\lambda$). Then the p^{th} order interference maxima on each side become absent.

wide slits are represented, the extreme members being A and B. The slit lengths are perpendicular to the plane of paper, widths are all equal to a and separations* are all equall to e.

The arrangement is illuminated by a monochromatic beam of light of wavelength λ falling normally from the left. At angles θ on the right we now have N waves, each of amplitude A_θ (see Eq. 14·5) and having successive phase differences $\Delta=(2\pi/\lambda)\, e \sin\theta$. We will find the resultant of these N waves by the phasor method.† In Fig. 14·11 (*b*) we draw equal lengths OP_1, $P_1P_2 \cdots P_{N-1}P_N$ (to represent the equal amplitudes A_θ) making equal inclinations Δ between successive ones (to represent equal phase differences). This gives a *regular polygon* and length OPn represents the resultant amplitude R_θ.

If C is the centre of the polygon, from simple geometry we find that angle subtended at C by each of OP_1, P_1P_2, $P_2P_3 \cdots$ is Δ. Hence

$$2.OC.\sin(\Delta/2)=OP_1=A_\theta,$$
$$2.OC.\sin(N\Delta/2)=OPn=R_\theta.$$

Eliminating OC, we get

$$R_\theta=A_\theta\,\frac{\sin(N\Delta/2)}{\sin(\Delta/2)}=A_0N\left(\frac{\sin\delta/2}{\delta/2}\right)\left(\frac{\sin N\Delta/2}{N\sin\Delta/2}\right)$$

substituting for A_θ from eq. (14·5). As $\theta\rightarrow 0$ (normal direction), both the bracketed terms tend to unity. Hence the intensity I_θ at angle θ, relative to the intensity I_0 *due to single slit* in the direction $\theta=0$, is given by

$$I_\theta=I_0\left[\left(\frac{\sin\delta/2}{\delta/2}\right)^2\right]\left[N^2\left(\frac{\sin N\Delta/2}{N\sin\Delta/2}\right)^2\right] \qquad \ldots(14{\cdot}15)$$

where we recall that

$$\delta=(2\pi/\lambda)\,a\sin\theta, \qquad \Delta=(2\pi/\lambda)e\sin\theta \qquad \ldots(14{\cdot}15\,a)$$

Check that for $N=2$, Eq. (14·15) reduces to Eq. (14·13).

---

* The *separation* is the distance between corresponding parts of successive slits, shown in detail in Fig. 14·9. If the opaque width is b, then $e=a+b$.

† An alternative method is this : Resultant disturbance y is given by

$$y=A_\theta\,[\cos\omega t+\cos(\omega t-\Delta)+\cos(\omega t-2\Delta)+\ldots\ldots N \text{ terms}]$$
$$=\text{real part of } A_\theta\, e^{i\omega t},\ [1+e^{i\Delta},+e^{2i\Delta}+\ldots\ldots N \text{ terms}]$$

After some working this leads to the form

$$y=A_\theta\frac{\sin N\Delta/2}{\sin\Delta/2}\cos(\omega t-\phi).$$

The amplitude of the expression is the desired R_θ.

In Eq. (14·15) the first bracket term expresses intensity distribution due to *diffraction* at the individual slit, and the second bracket term arises from *interference* between the N diffracted waves.

Diffraction spread. The first minimum due to diffraction occurs at $\theta=\pm\sin^{-1}(\lambda/a)$ as already discussed in § 14·2.

(b) Principal interference maxima. The interference term becomes N^2 when $\sin \triangle/2=0$ and this is the *maximum* value it can take.* The maxima thus occur when $\triangle/2=n\pi$, which leads to

$$\boxed{e \sin \theta=n\lambda}, \quad n=0, 1, 2, \cdots \qquad \ldots(14·16)$$

This gives the *principal maxima* ; n is called the *order* of the interference maximum. Condition (14·16) simply says that maxima occur when waves from successive elements reach with a path difference of a whole number of λ.

The intensity at the centre of any principal maximum is, from Eq. (14·15), given by N^2 times that due to a single slit at that angle.

(c) The minima. The interference term in (14·15) becomes *minimum* (=0) whenever the numerator is zero (excluding cases where the denominator is also zero), *i.e.*, when $N\triangle/2=r\pi$, which leads to

$$\boxed{N e \sin \theta=r\lambda} \qquad \ldots(14·17)$$

where r is any integer excluding $O, N, 2N, \ldots nN \ldots$

Our interest is in the minima nearest to the n^{th} order maximum for which $r=nN\pm1$. If this occurs at angle $\theta\pm\delta\theta$,

$$e \sin(\theta\pm\delta\theta)=\left(\frac{nN\pm1}{N}\right)\lambda=n\lambda\pm\frac{\lambda}{N}$$

Taking difference with $e \sin \theta=n\lambda$, we get

$$e \cos \theta.\delta\theta=\pm\,\lambda/N$$

or

$$\boxed{|\,\delta\theta\,|=\frac{\lambda}{Ne \cos \theta}} \qquad \ldots(14·18)$$

This is called *angular half-width* of the n^{th} order maximum. Note that it varies in proportion to $1/N$; *the interference maxima become sharper as the number of openings is increased.*

---

* $\lim_{p\to n\pi}\left(\frac{\sin Np}{N \sin p}\right)=1.$

Secondary maxima. It is left as an exercise to the reader to show that between two principal maxima (say, given by $r=nN$ and $\overline{n+1}N$) there will be $N-1$ minima and $N-2$ secondary maxima; also that for large N the secondary maxima nearest to a principal maximum will have relative intensities approximately

$$(3\pi/2)^{-2},\ (5\pi/2)^{-2},\ (7\pi/2)^{-8},\ \text{etc.}$$

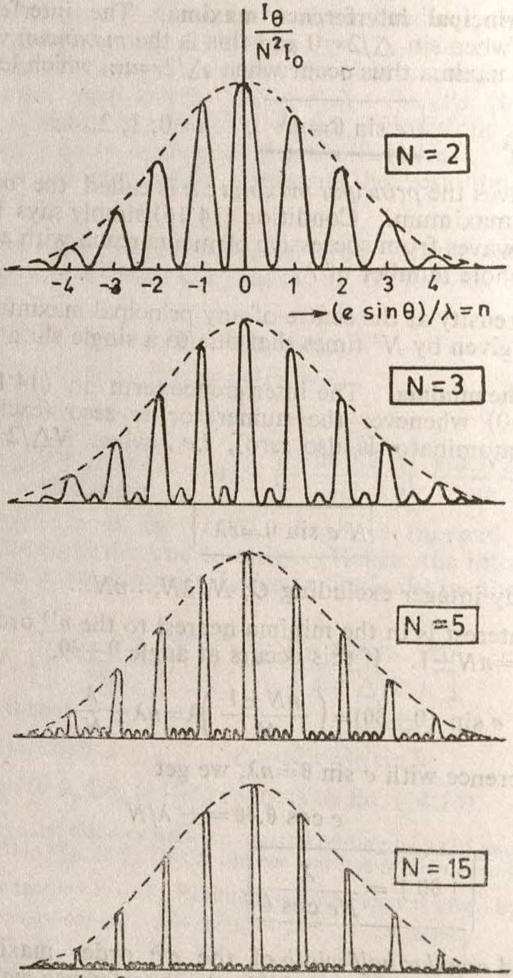


Fig. 14.12.*

Intensity distribution in Fraunhofer diffraction due to N parallel equidistant slits. The peaks narrow down as N increases

---

[*] The Secondary maxima are shown exaggerated in fig. 14·12. Also, the horizontal scale is proportional to sin θ, not θ. On θ scale the maxima would not be equidistant.

Fig. 14·12 shows the intensity distribution for $N=2, 3, 5, 15$ slits, keeping n and e the same. We have taken the case of $e \approx 4{\cdot}5a$, hence within the central diffraction maximum (shown by the dotted curve) four principal interference maxima are seen on either side of the zero order. For two slits ($N=2$), there are no secondary maxima in between, and the curve is like the bottom curve of Fig. 14·9. For $N=5$, there are 3 *secondary maxima* between the principal maxima. For $N=15$ the secondary minima are not distinct in the figure, but the principal maxima are sharper.

Example 14·4. A set of 10 parallel equidistant slits of width 0·50 cm each and spacing $e=1{\cdot}90$ cm are used to study Fraunhofer diffraction of waves of $\lambda=0{\cdot}60$ cm falling normally on the plane of the slits. Deduce (i) angular position of the interference maxima, (ii) half-widths of these maxima, (iii) the effect of covering up 5 slits from one end, (iv) the effect of covering up alternate slits.

Solution. The data given are :

$$a=0{\cdot}50 \text{ cm}, \qquad e=1{\cdot}90 \text{ cm}, \qquad \lambda=0{\cdot}60 \text{ cm}, \qquad N=10.$$

(*i*) Eq. (14·16) gives the interference maxima :

$$\sin\theta=\frac{0{\cdot}60}{1{\cdot}90}, \quad \frac{2\times 0{\cdot}60}{1{\cdot}90}, \quad \frac{3\times 0{\cdot}60}{1{\cdot}90}$$

$$\therefore \quad \theta=\mathbf{18^\circ 24'}, \quad \mathbf{39^\circ 18'}, \quad \mathbf{71^\circ 30'}$$

(*ii*) Eq. (14·18) gives the half-width

$$\delta\theta=\frac{0{\cdot}60}{10\times 1{\cdot}90\cos\theta}$$

$$=\mathbf{{\cdot}033}, \quad \mathbf{{\cdot}041}, \quad \mathbf{{\cdot}100} \text{ radian.}$$

(*iii*) This step reduces N to 5, keeping e unchanged. Hence θ remains unaltered, $\delta\theta$ becomes doubled.

(*iv*) The step reduces N to 5, and doubles e. Hence the θ deduced above now correspond to II, IV and VI order maxima, their half widths remain unaltered. [Positions and half-widths of I, III, V orders may be deduced too.] ⊙

14·8. Plane Diffraction* Grating.

The theory above would show that N parallel and equidistant slits, with large N, should produce dispersion with sharp maxima for different λ values

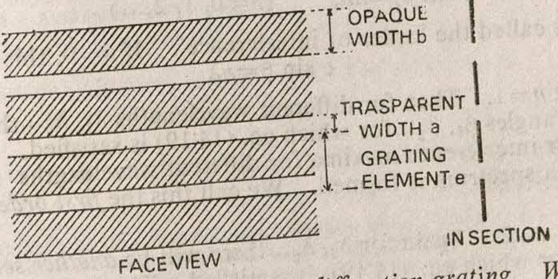


Fig. 14.13. Magnified view of a diffraction grating. Width a is usually $\sim 10^{-4}$ *cm and e is usually* $\sim 3a$

---

* The term *diffraction*' is not appropriate, but has become current. The spectra are produced by *interference* of N wavelets ; diffraction only serves to send light in the required directions.

However, for visible light the spacing e has to be of the order of $\sim 10^{-4}$ cm (a few times λ). Producing alternate transparent and opaque widths at equal spacings was first achieved by Rowland, and the device is called Rowland grating. It is now one of the most important tools in spectroscopy.

Diffraction gratings are of two types : (*a*) Transmission type and (*b*) Reflection type. We shall discuss the first in some detail, most results being true in the second also.

(*i*) **Description.** The transmission type diffraction grating is a plane plate of glass on which a very large number (N) of parallel rulings of width b each are drawn, with clear spaces of width a each between them. The ruled widths become opaque and clear spaces are left transparent. The width $a+b=e$ is called the grating element. In gratings used for visible light a is of the order of 10^{-4} cm. Usually, e is about 3 or 4 times a. Fig. 14·13 shows a highly magnified and simplified* view of a small part of the grating in face view and in section.

(*ii*) **Formation of spectrum by the grating.** Let light fall on the grating as a parallel beam at angle α from the normal (see Fig. 14·14)· Let us observe the diffracted light at angle β from the normal on the other, side. The path difference between waves from successive elements is, from Fig. 14·14,

$$e\,(\sin\alpha+\sin\beta).$$

For any wavelength λ, we get a maximum in the direction β if this path difference satisfies

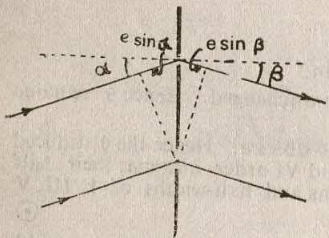


Fig. 14·14.
Path difference for waves from successive elements

$$e\,(\sin\alpha+\sin\beta)=n\lambda, \qquad (n=0, 1, 2, \ldots) \qquad \ldots(14\text{·}19\text{ a})$$

Here n is called the 'order of interference'. If $\alpha=0$, this reduces to

$$e\sin\beta=n\lambda \qquad \ldots(14\text{·}19\text{ b})$$

Put $n=1$. Then for different wavelengths λ_1, λ_2...there will be different angles β_1, β_2...for which eq. (14·19) is satisfied. Thus, the first order interference maxima for different wavelengths are separated, *i.e* , a spectrum is formed. We call this the *first order spectrum* ($n=1$).

Put $n=2$. Again for λ_1, λ_2...there will be *another set* of angles β_1, β_2... for which eq. (14·19) is satisfied. This gives the *second order spectrum*. And so on.

(*iii*) **Dispersion.** Let the difference of two wavelengths be $\Delta\lambda$ and let their nth order maxima be formed at angular separation

---

* One need not have a fully transparant and b fully opaque. Any variation of transmittance repeated at intervals e each will produce the spectrum given by Eq. (14·19). The first bracket term in Eq. (14·15) will differ, but not the second one.

Note that for $n=0$ all wavelengths satisfy (14·19) at one single angle β. This gives the *zero order 'spectrum'*. But since there is no separation of different wavelengths, this zero order image is of no use. It only gives the overall colour of the source.

(*iii*) *Dispersion*. Let the difference of two wavelengths be $\Delta\lambda$ and let their *n*th order maxima be formed at angular separation $\Delta\beta$. Then, by definition $\Delta\beta/\Delta\lambda$ is called *angular dispersion*. From eq. (14·19), by differentiation we get (since α is constant)

$$e \cos \beta . \frac{\Delta\beta}{\Delta\lambda}=n$$

or

$$\text{dispersion}=\frac{\Delta\beta}{\Delta\lambda}=\frac{n}{e \cos \beta} . \quad \text{...(14·20)}$$

Thus the dispersion is higher in higher orders due to (*i*) increase of *n*, and (*ii*) decrease of cos β. (See Example 14·7).

The significance of having *N* large, may be seen from eq. 14·18. We shall discuss this matter in detail in § 15·4.

(*iv*) **Different settings in spectrographs.** A typical spectrograph using a grating involves a collimator, the grating, and a telescope or camera. The slit of the collimator is kept parallel to length of rulings on the grating. The angles of incidence and observation (α and β) are set up in many different ways, two of which are—

(*a*) **Normal incidence.** Here $\alpha=0$, and various orders of spectra occur on both sides symmetrically according to eq. (14·19 b)

(*b*) **Normal observation.** In this case α is so set that the desired order of spectrum for the given region of λ appears *near* $\beta=0$. The dispersion (given by eq. 14·20) then remains almost uniform over the observed spectum. Such spectrum is called a '*normal spectrum*.'

Example 14·5. A plane transmission grating has $e=2\times10^{-4}$ cm. Deduce the positions of first, second and third order maxima for (i) $\lambda=5\times10^{-5}$ cm, (ii) $\lambda=8\times10^{-5}$ cm. Assume normal incidence.

Solution. Using eq. (14·19*b*),

(*i*) For $\lambda=5\times10^{-5}$ cm, $\sin\beta=\frac{n\lambda}{e}$ 0·25, 0·50, 0·75 (check)

$\therefore$ $\beta=14°30', 30°, 48°36'$.

(*ii*) For $\lambda=8\times10^{-5}$ cm, $\sin\beta=\frac{n\lambda}{e}=0·40, 0.80, (1·20)$

$\therefore$ $\beta=23°36', 53°6'$. (No third order) ⊙

Example 14·6. A plane diffraction grating has $e=15\times10^{-4}$ cm. Calculate the position of the third order maximum for $\lambda=2·4\times10^{-4}$ cm. (This wavelength is in infra-red region). For which other longer wavelengths do the maxima fall in the same position ? Assume normal incidence.

Solution. $\sin\beta=\frac{n\lambda}{e}=\frac{3\times2·4\times10^{-4}\text{ cm}}{1·5\times10^{-4}\text{ cm}}=0·48 \quad \therefore\ \beta=29°$

For *longer* wavelengths giving maxima at the same angle, we have to consider smaller *n*, that is $n=2$ and $n=1$. If the wavelengths are λ' and λ'' we have

$3\lambda=2\lambda'=1\lambda'' \quad \Rightarrow \quad \lambda'=3·6\times10^{-4}$ cm, $\lambda''=7·2\times10^{-4}$ cm ⊙

Example 14·7. A plane diffraction grating of $e=17\times10^{-5}$ cm is used to photograph a spectrum using normal incidence. Calculate angular dispersion in the region of the different orders of the spectrum of $\lambda=5{\cdot}9\times10^{-5}$ cm. If the camera has focal length 25 cm, calculate the linear dispersion in the spectrograph, and also the separation between spectral lines 5890A° and 5896 A° in the second order.

Solution. Eqs. (14·19*b*) and (14·20) are useful here. For the three orders—

$$\beta=\sin^{-1}\left(\frac{n\lambda}{e}\right)=\sin^{-1}(\cdot311),\ \sin^{-1}(\cdot622),\ \sin^{-1}(\cdot933)$$

$$=18^\circ6',\qquad 38^\circ30',\qquad 68^\circ54'.$$

Corresponding values of cos β are 0·951, 0·783, 0·360

$$\frac{\triangle\beta}{\triangle\lambda}=\frac{n}{e\cos\beta}=5{\cdot}7\times10^{3},\ 13{\cdot}4\times10^{3},\ 43{\cdot}9\times10^{3}\ \text{rad/cm.}$$

$$=5{\cdot}7\times10^{-5},\ 13{\cdot}4\times10^{-5},\ 43{\cdot}9\times10^{-5}\ \text{rad}/A^\circ.$$

Linear dispersion is $f.\ \triangle\beta/\triangle\lambda$. Calling it $\triangle x/\triangle\lambda$,

$$\frac{\triangle x}{\triangle\lambda}\ \text{for second order}=25\times13{\cdot}4\times10^{-5}\ \text{cm/A}^\circ$$

$$=3{\cdot}35\times10^{-3}\ \text{cm}/A^\circ$$

For $\triangle\lambda=6\text{A}^\circ$ we get

$$\triangle x=\frac{\triangle x}{\triangle\lambda}.\ \triangle\lambda=6\times3{\cdot}35\times10^{-3}=2{\cdot}04\times10^{-2}\ \text{cm.}$$ ⊙

14·9. Plane Reflection Grating. Consider an optically plane surface with its surface made reflecting by silvering or aluminizing. If a large number (N) of parallel straight lines are scratched on it, each of width b and leaving equal widths a clear between them, we get a plane *reflection* grating.

The reflection grating is used to obtain spectra just like the plane transmission grating, the theory and action being the same. Reflection grating has the particular advantage that it can be used also in the ultra-violet and infra-red regions where most materials become opaque. Hence it is very widely used.

14·10. Blazed Grating. While higher order spectra are desirable for good dispersion, the intensity is maximum in the zero order and falls as n increases (see Fig. 14·12). Blazed grating is designed to shift the intensity maximum towards a desired high order spectrum. Grooves will inclined reflecting faces are made in the grating surface (Fig. 14·15). The inclination γ is called the *blaze angle.*

If light is incident along normal AO to the grating surface *BB*. the *central diffraction maximum* now falls not along OA, but at on angle 2γ from OA. The spectrum seen at angle β from OA corresponds to

$$e\sin\beta=n\lambda$$

The important thing is that if β is in the neighbourhood of 2γ, then the spectrum has large intensity, despite n being high. If one desires to choose a grating blazed for a given order n for a given region λ then one calculates β and sets $\gamma=\frac{1}{2}\beta$ as the blaze angle.

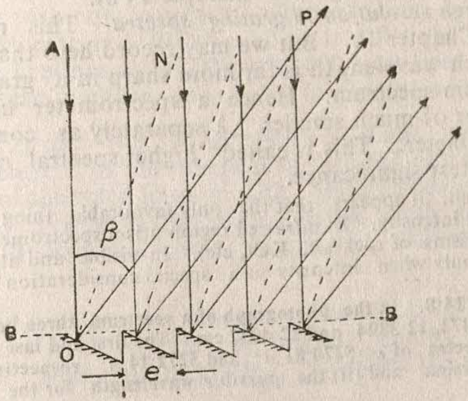


Fig. 14·15.

Blazed diffraction grating

14·11. Some Special Features of Grating Spectra. It is useful to discuss some special features of grating spectra, particularly in contrast with prism spectra.

(*a*) *Intensity consideration in grating spectra.* When a spectrum is obtained with a prism, all light for each wavelength is collected in the corresponding single spectral image. In contrast, with the grating the spectrum is very weak for two reasons.

(*i*) Firstly, out of each width e of the beam, only a fraction a/e is used.

(*ii*) Secondly, this light goes into several 'orders' of spectra. Of these, the 'zero order' which takes the largest amount of light, is wasted*, since the maxima for all the wavelengths coincide in this case.

(*b*) *Rational and irrational spectra.* Prism spectrum depends on the variation of refractive index μ with wavelength λ, which varies from one glass to another considerably. And further, the deviation ϕ depends upon μ and angle of the prism in a complicated way. So the spectrum is called '*irrational*' (not following any well-defined law between ϕ and λ). In contrast, the grating spectum has a definite relation between β and λ eq. (14·19). The spectrum is therefore called '*rational*'.

(*c*) *Overlapping spectra.* In grating spectra there is overlapping of the different orders of spectra for different wavelengths. For example in the position of the second order for 6500*A*°, we also get the third order of 4333$A°$, fourth order of 3250$A°$, fifth order of 2600$A°$... The overlapping causes troubles sometimes. But it can also be used to advantage (see Example 14·8).

---

* Except in blazed reflection gratings. (Fig. 14·15).

(*d*) *High resolution in grating spectra.* This matter will be taken up in Chapter 15. But we may record here that the spectral image for each wavelength is far more sharp in a grating spectrum than in a prism spectrum. Hence a spectrometer using a grating can show lines of much smaller $\triangle\lambda$ separately as compared with a prism spectrometer. This is called 'higher spectral *resolution*', and is of the greatest significance.

All in all, it appears that the only favourable thing about a prism spectrum is high intensity. In infra-red region prism spectrometers are still used quite widely (prisms of rocksalt, KCl, etc). In visible and ultraviolet regions they are used only when intensity is a special consideration (*e.g.* in Raman effect).

Example 14·8. In the photograph of a spectrum three lines are observed at positions 12·3472, 12·3894 and 12·4037 cm. The first and last are known to be second order spectra of λ=5270·81 $A°$ and 5272·14 $A°$ respectively. Deduce (i) the linear dispersion, and (ii) the *possible* wavelength for the spectral line at 12·3894 cm.

Solution. Linear dispersion$=\dfrac{\triangle x}{\triangle\lambda}=\dfrac{0{\cdot}0565 \text{ cm}}{1{\cdot}33A°}=0{\cdot}0425$ cm per $A°$

Let the third line also be the *second* order spectrum of some λ. For small regions of λ we can use linear dispersion. Hence

$$\lambda=5270{\cdot}81\,A+\frac{(12{\cdot}3894-12{\cdot}3472)\text{ cm}}{0{\cdot}0425\text{ cm}/A}=\mathbf{5271{\cdot}80\ A}$$

But this line could also be the first order spectrum of 2 × 5271·80, or third order spectrum of $\frac{2}{3}\times 5271{\cdot}80$ A° etc. Thus the *possible* λ values are

10543·60 A°, 5271·80 A°, 3541.53 A°,... ⊙

14·12. Concave Grating. Use of a plane grating needs a collimating and a focussing system. Rowland discovered that if a grating is ruled on a *concave* spherial reflecting surface, then under some conditions it produces focussed spectra by itself.

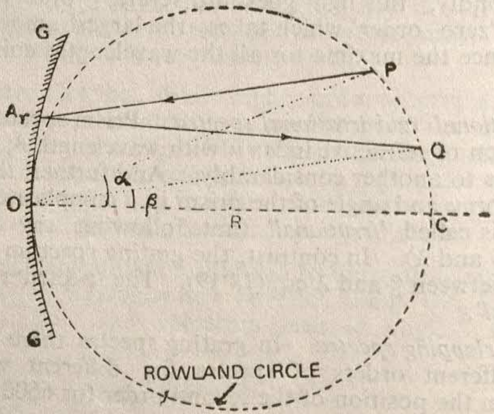


Fig. 14·16.

The Rowland grating and Rowland circle

Consider a spherical front-silvered mirror GOG (Fig. 14.16). Let a large number of parallel lines of narrow widths b be drawn on this surface with equal spaces of width a left in between. This forms a concave grating. Width $a+b=e$ is called the grating element. The usual rulings are 10,000 to 30,000 lines per inch. Their distances are *equal not along the arc but in projection on the chord.*

Rowland circle. Fig. 14·16 shows a plane section passing through the centre C of the spherical surface GG and perpendicular to the length of the ruled lines on GG. With CO *as diameter* (not as radius) draw a circle in this plane (dotted circle in Fig. 14·16). This is called the Rowland circle. In practice OC may be a few metres.

Focussing and spectrun formation. Let P and Q be two points on the Rowland circle specified by angles α and β. Let A_r be a point on the r^{th} ruling on the grating. Consider the path

$$d_r = PA_r + A_rQ$$

Rowland showed* that for P,Q located on this circle there are two important properties—

(i) As we vary the position A_r *on a given ruling*, the path-length d_r does not change.

(ii) As we shift the position of A to a *neighbouring ruling*, d_r changes by an amount p which is independent of r and is given by

$$| d_r - d_{r+1} | \equiv p = e(\sin\alpha + \sin\beta) \qquad ...(14·22)$$

The first of these properties means that the secondary wavelets from all parts of a given ruling reach Q in the same phase. The second property means that wavelets diffracted by successive rulings reach Q with *equal path differences p*.

If this path difference is a whole number of a given λ, then a bright 'image' of P for wavelength λ is formed at Q. The condition is

$$e(\sin\alpha + \sin\beta) = n\lambda, \qquad (n=0,1,2,\dots) \qquad ...(14·23)$$

For a given position of P(*i.e.*, given x), this is satisfied for different wavelengths at different angles β. Thus spectra are formed. Notice that no collimating or focussing device is needed.

Theory of the Rowland Grating. In fig 14.17 the curved line is a section of the grating surface with C as its centre. P and Q are the source point and observation point in this section. We choose OC as x-axis, the y-axis tangent to the grating in this plane, and z-axis normal to the figure at O.

With R as radius of curvature of the grating, the co-ordinates of C are $(R, 0, 0)$. P and Q are general points in x-y plane. On the grating surface we choose a general point $A(x,y,z)$, and then search for the kind of ruling on the grating and the relative locations of P, Q so that waves from P may be focussed at Q for a given λ.

---

* See below.

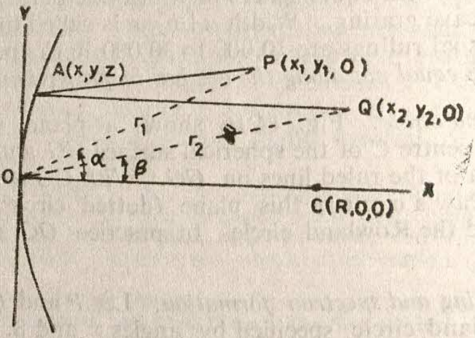


Fig. 14·17.

The focussing condition in concave grating

Let the ruled area on the grating in y and z directions be so limited that (as compared to distances R, r_1,r_2) we can neglect (*i*) second and higher powers of z, and (*ii*) fourth and higher powers of y. The x coordinate for A is already of second order relative to y and z, hence x^2 terms can be neglected. In the following calculations we shall keep these in mind, as also the facts that $r_1^2=x_1^2+y_1^2$ and $r_2^2=x_2^2+y_2^2$. Now—

$$(PA)^2=(x-x_1)^2+(y-y_1)^2+z^2=r_1^2+y^2-2xx_1-2yy_1 \qquad \ldots(14{\cdot}24)$$

But $$(AC)^2=R^2=(x-R)^2+y^2+z^2\approx R^2-2xR+y^2,$$

$$\therefore\ 2x=y^2/R.$$

Substituting in eq. (14·24) and arranging in powers of y

$$PA=\left[r_1^2-2yy_1+\left(1-\frac{x_1}{R}\right)y^2\right]^{\frac{1}{2}}$$

$$=r_1[1+fy+gy^2]^{1/2} \qquad \ldots(14{\cdot}25)$$

where $$-\frac{2y_1}{r_1^2}=f \text{ and } \frac{1}{r_1^2}\left(1-\frac{x_1}{R}\right)=g. \qquad \ldots(14{\cdot}26)$$

Since $fy+gy^2 \ll 1$, we expand (14·25) to terms upto y^2

$$PA=r_1[1+\tfrac{1}{2}(fy+gy^2)-\frac{1}{8}(fy+gy^2)^2+\frac{1}{16}(fy+gy^2)^3]$$

$$=r_1[1+\tfrac{1}{2}fy+\tfrac{1}{2}y^2(g-\frac{1}{4}f^2)-\frac{1}{4}fy^3(g-\frac{1}{4}f^2)]. \qquad \ldots(14{\cdot}27)$$

We note from eq. (14·26) that f and g are related with the location of point P. If we locate P to satisfy the condition $(g-\frac{1}{4}f^2)=0$, then both y^2 and y^3 terms in (14·27) drop out. The condition required is

$$g-\tfrac{1}{4}f^2=0, \quad \text{or} \quad 1-\frac{x^1}{R}-\frac{y_1^2}{r_1^2}=0. \qquad \ldots(14{\cdot}28)$$

This is the equation to Rowland circle, which is the circle on OC as diameter. The condition tells us that *P should lie on the Rowland circle* to make the coefficients of y^2 and y^3 vanish in eq. (14·27), which then gives

$$PA=r_1\left(1-\frac{y_1 y}{r_1^2}\right)=r_1-\frac{y_1 y}{r_1}=r_1-y\sin\alpha.$$

If Q also lies on Rowland circle.

$$AQ=r_2-y\sin\beta$$

Hence total path $PA+AQ$ is given by

$$d=PA+AQ=(r_1+r_2)-y(\sin\alpha+\sin\beta) \qquad \text{...(14·29)}$$

To keep d constant as A moves along a given ruling, we require y=constant. *A given ruling on the grating must therefore be perpendicular to the plane of the Rowland circle* (*i.e., along the z-axis* in our figure).

Let *successive* rulings have a separation $\triangle y=e$. Then as A steps to successive rulings, path d changes by ...(14·30)

$$p=e\,(\sin\alpha+\sin\beta).$$

same as eq. (14·22). To keep p constant for all pairs of rulings we must have Δy constant *i.e., successive rulings must be equidistant along the chord.*

14·13. Different Mountings of a Concave Grating.* In using a concave grating it is necessary to ensure that the slit and photographic plate are on the Rowland circle. One of three different arrangements, called mountings, are used to secure thes econditions.

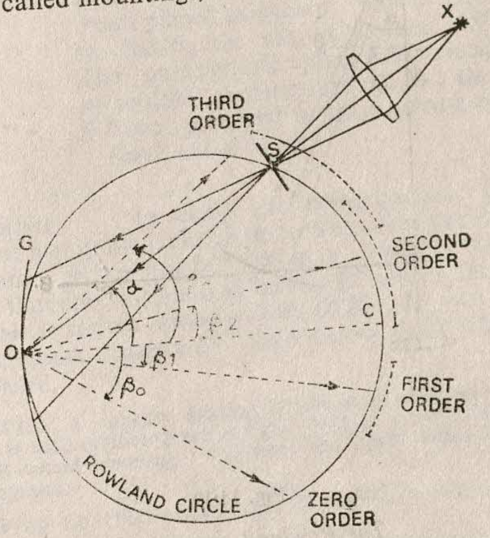


Fig. 14·18.

The Runge Paschen mounting

---

* Rowland circle mountings only are described here. There are a few other mountings also.

Runge-Paschen Mounting. In this a horizontal circle of cement and concrete is made in the laboratory with diameter equal to radius of curvature of the concave grating. A steel railing is set on this. The grating G (Fig. 14·18) is fixed at one location on the circle with its rulings normal to the plane of the circle. Slit S is mounted (Fig. 14·18) on the railing at a suitable angle $SOC=\alpha$, with the length set normal to the Rowland circle. Spectra of various orders for all wavelengths in the source are then formed as shown schematically in Fig. 14·18. Either a photographic film is spread into a groove all along the Rowland circle, or photographic plates in several parts are set along the circle to record *simultaneously several orders of spectra.*

Rowland Mounting. This mounting (Fig. 14·19) has two fixed railings SA, SB on which the ends of a rigid bar GP can slide on rollers. The bar has length equal to R of the grating, and two positions (not two different bars) GP and $G'P'$ are shown in Fig. 14·19.

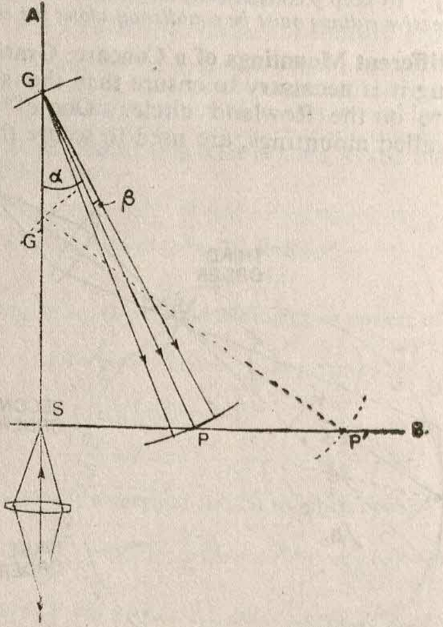


Fig. 14·19.

The Rowland mounting

Now the slit is placed at the junction S, with its length normal to plane ASB. The grating G is mounted on the bar at one end such that its axis coincides with bar length GP and the length of rulings is normal to plane ASB. With this setting, as the bar GP slides the Rowland circle always passes through P and also throuth S. The photographic plate or film is moundet no the bar at the and P.

As the bar GP is slided, the angle α changes. The angle β is always *in the neighhourhood of zero* (it is zero at centre of the plate P). The dispersion $n/(e \cos \beta)$ therefore remains invariant $(=n/e)$.

Eagle's Mounting. In this arrangement, the grating GG (Fig. 14·20) is mounted on a stand in an ordinary optical bench AB. The garting can be moved along AB and can be turned about a vertical axis. Slit S and a small right angled prism P are placed as shown. This is equivalent to having the slit at S' where $PS'=PS'$. The photographic plate EF is placed with its middle passing through S'; the bellws B B are provided to allow setting of plates F at a suitable tilt.

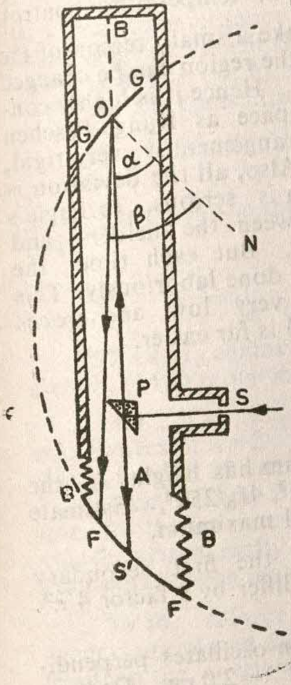


Fig. 14·20.
Eagle's Mounting

In using the arrangement, first the grating GG is set at a suitable inclination α. Then it is moved along AB so as to get some spectral line in focus at S'. Finally, the tilt of FF is adjusted so that spectral lines all along FF are in fair focus. The Rowland circle is shown dotted in Fig. 14·20. It can be seen that if tilt α is increased, distance OS' must be decreased and tilt of FF must be increased.

Example 14·12. A concave grating of 60000 lines/cm and radius of curvature 5·0 metre is used to photograph sodium D-lines (5890,5896 A) in the second order with Rowland mounting. Calculate the liuear separation of the doublet on the photograph.

Solution. From Eq. (14·31), the angular dispersion is given by

$$\frac{\delta\beta}{\delta\lambda}=\frac{n}{e\cos\beta}\approx\frac{n}{e}\ \text{(since } \beta\approx 0 \text{ in Rowland mounting.)}$$

Linear separation $\delta x=R\delta\beta$. Hence from the given data $n=2$, $e=(1/6000)$ cm, $\delta\lambda=6\times10^{-8}$ cm and $R=500$ cm.

$$\delta x=\frac{nR}{e}\,\delta\lambda=2\times500\text{ cm}\times6000\text{ cm}^{-1}\times6\times10^{-8}\text{ cm}=0{\cdot}36\text{ cm}.$$

Comparison. When extensive regions of spectra are to be studied simultaneously, the Runge-Paschen mounting alone is useful. With rare materials this is particularly important. But this mounting needs a large room, and involves problems of temperature control.

The Rowland and Eagle mountings take a small region of the spectum at a time. In Rowland mounting the region can be changed automatically and the spectrum is 'normal'. Hence it is highly convenient. But it occupies almost the same space as Runge-Paschen mounting and unless the mechanical arrangement is very rigid, finer focussing has to adjusted each time. Also, all the deviation is in the incident beam, hence astigmatism is serious. In Eagle's mounting deviations are equally divided between the incident and diffracted beams, hence astigmatism is least. But each time the region of study changes, the setting has to be done laboriously. This labour is compensated, because the cost is very low and spaces occupied is small so that temperature control is far easier.

PROBLEMS

14·1. In Fig. 14·3 if the central maximum has height I_o, the other maxima have heights roughly $4I_o/9\pi^2$, $4I_0/25\pi^2$,...Estimate what fraction of the total light is in the central maximum.

14·2. The expressions for intensity in the first secondary maximum as given in eqs. (14·3*b*) and (14·10) differ by a factor $4/\pi^2$. Explain the physical reason for this.

14·3. A rectangular plate 15 cm by 4·0 cm oscillates perpendicular to itself to generate longitudinal waves of $\lambda=2·0$ cm. Deduce the angular spread of the waves from the direction of normal to the plate.

14·4. Show that if light falls on a slit at angle of incidence β and observations are made at angle θ, the minima occur when

$$a\,(\sin\beta-\sin\theta)=n\lambda$$

where a is slit width, and β and θ are both measured clockwise from the normal.

14·5. Plane waves of $\lambda=2·5$ cm are passed through a slit whose width varies from 10·0 cm to 2·0 cm. Discuss the variations in the angular spread of the waves on the other side.

14·6. The slit in a collimator has width 2×10^{-3} cm, the lens having focal length 20 cm and aperture diameter 3 cm. Deduce angular spread of the beam due to the width of the slit, and due to diffraction at the lens (taking $\lambda=6\times10^{-5}$ cm).

14·7. Plane waves of $\lambda=5\times10^{-5}$ cm fall normally on a slit of width 0·20 mm. Deduce the total width of the central maximum observed on a screen placed 2·0 metres away, using a lens to focus the waves.

14·8. Light of wavelength λ falls normally on a slit of width a, whose transmittance is so modulated that at distance x from the middle-line of the slit the amplitude of the transmitted wave per unit width is $A_0 \cos(\pi x/b)$, with $b \gg a$. Show that the intensity distribution of the transmitted beam is

$$\frac{I_\theta}{I_\theta} = \left[\frac{\sin(q+p)}{2(q+p)} + \frac{\sin(q-p)}{2(q-p)} \right]^2 ; p = \frac{\pi a \sin\theta}{\lambda}, \quad q = \frac{\pi a}{2b}.$$

Show that for $b \to \infty$ this reduces to eq. (14·5), and for $b = a$ it becomes

$$\frac{I_\theta}{I_0} = \left[\frac{\pi/2}{(\pi/2)^2 - p^2} \right]^2 \cos^2 p.$$

Plot I_θ / I_0 against θ for the case $a = 2\lambda$. [Hint : Use Fig. 14·7 and eq. (14·7) replacing A by $A_0 \cos \pi x/b$.].

14·9. Two long slits are placed parallel and illuminated with microwaves of $\lambda = 2·5$ cm falling normally. The spacing e of the slits is fixed at 20 cm, but opening a of each slit is varied in steps 3 cm, 5 cm, 10 cm, 20 cm. Discuss and draw the intensity distributions for the four cases. (Use the method of Fig. 14·9. With $a = 20$ cm, the system really becomes a *single* slit of width 40 cm. Check if your conclusions match this).

14·10. A large number of oscillators, controlled by a common agency, are placed along a line equidistantly. The spacing is much less than the wavelength λ of their radiations. Discuss the space distribution of their main beam when total length of the line of oscillators is 10 λ and all oscillators are in the same phase. Show also that if the line is made n-fold the central beam intensity increases n^2 fold.

14·11. Compare Figs. 13·12 and 14·6 and explain why the phasor diagram forms a spiral shape in one case and circular shape in the other.

14·12. Eq. (14·15) shows I_θ is proportional to N^2. But doubling the number of slits would only double the available energy. How do you reconcile that I_θ is not proportional to N, but to N^2 ?

14 13 A plane diffraction grating has 5000 lines per cm. How many orders of spectra will it show for $\lambda = 5 \times 10^{-5}$ cm using (*i*) normal incidence, (*ii*) normal observation, (*iii*) minimum deviation ? Also deduce which wavelengths in the range 3500 *A* to 8000 *A* will overlap with the third order for $\lambda = 5 \times 10^{-5}$ cm.

14·14. A plane diffraction grating may be set for viewing a spectrum under 'minimum deviation' condition. If the minimum deviation is ϕ, show that $2e \sin \frac{1}{2} \phi = n\lambda$. What may be the advantage in minimum deviation setting ?

14·15. Wire-gratings are used in far infra-red. nI one caes 1000 wires of diameter 0·2 mm each are stretchd parallel to noe another leaving spacings of 0·05 mm each. Deduce the wavelength for which the third order maximum falls at 30°, incident light falling normally.

14·16. A spectrum is seen with a plane diffraction garing. How will this beaffected if (*a*) the grating is tis moved in own place, (*b*) the grating is turned in its own plane, (*c*) the grating is moved perpendicular to its own plane, (*d*) the grating is turned abut an axis parallel to the lines on it ?

14·17. A blazed grating has 6000 lines per cm and is dlazed for the order third order of $\lambda = 2500$ A°. Calculate the blaze angle.

14·18. In a concave grating with Rowland mounting the spectrum shows four lines whose positions on the photographic plate are 0·2174 mm, 0·1493 mm, 0·0897 mm, and 0·0127 mm. If the wavelengths of the first and last lines among the four are 5134·6 *A*° and 5172·3*A*°, calcuate the wavelengths of the other two lines.

14·19. A concave grating with 6000 lines per cm and radius of curvature 2·4 meters is used to study the spectrum near 4000*A*° in the second order. Deduce the linear dispersion if the mounting used is Rowland's.

14·20. "The theory of plane diffraction grating does *not* depend on assuming that widths *a* are fully transmitting and widths *b* are fully opaque. It only requires that there are variations in the transmitting power with repetitive spacing *e*." Comment on this

15

Resolving Power of Optical Instruments

15·1. **'Geometrical' and 'Spectral' Resolution.** The capacity of an optical instrument to show two close things separately is called '*resolution*'. If a fine wire gauge is placed several metres away, our naked eye cannot see the details. If a source of light emits light of two wavelengths, whose difference is very small, an ordinary spectroscope cannot show the two spectral lines separately. In both cases we say that the details are 'not resolved' by the instrument used. A better instrument may 'resolve' the details.

Throughout this Chapter we shall assume that the instruments, are free from all aberrations—*i.e.*, they are designed to give the ideal image, *as far as nature permits*. The wave character of light shows that this *ideal image has a finite spread*. The effect of this spread on the resolving capacity of an instrument will be studied in this chapter.

We must distinguish between cases where geometrical positions between two near objects are to be 'resolved' (*geometrical resolution*) and cases where small differences of wavelengths of light in a given source are to be 'resolved' (*spectral resolution*). A grating or a prism gives us spectral resolution. The definition is :

$$\textit{Spectral resolving power}=\frac{\lambda}{\triangle\lambda m} \qquad \ldots(15{\cdot}1)$$

where $\triangle\lambda m$ is the smallest difference of wavelength which can be resolved by the instrument at the wavelength λ.

A telescope gives us geometrical resolution between two *distant objects* subtending a very small *angle*. We define

$$\textit{Resolving limit of a teleseope}=\theta_m, \qquad \ldots(15{\cdot}2)$$

where θm is the smallest angle resolved by the telescope.

A microscope also gives geometrical resolution, but here the linear *distance* between two *close objects* is to be resolved. We define

Resolving limit of a microscope$=x_m$, ...(15·3)

where x_m is the smallest distance resolved by the microscope.

It may be noted that θ_m and x_m will be smaller for instruments capable of showing finer details. In that sense, it is better to use the term *'resolving limit'*, rather than resolving power, in the last two definitions.

For the unaided human eye, θ_m is of the order of 1 minute of arc ($\sim 1/3500$ radian) and x_m is of the order of 0·1 mm.

Example 15·1. Deduce the resolving power of an instrument which can show lines of $\Delta\lambda=0{\cdot}05\ A$ at $\lambda=6000\ A$ separately.

Solution. By definition,

$$RP=\frac{6000\ A}{0.05\ A}=1{,}20{,}000.$$

The instrument should have RP above 1,20,000. ⊙

Example 15·2. Deduce the resolving power of a camera which can photograph details of 0·2 cm size from a distance of 100 metres.

Solution. By definition,

$$RP=\theta_{min}=\frac{0{\cdot}2\ \text{cm}}{10{,}000\ \text{cm}}=2\times10^{-5}\ \text{radian}.$$

The camera should have RP better than 2×10^{-5} radian. ⊙

15·2. Distinction between 'Magnification' and 'Resolution'. Let us consider geometrical resolution by a telescope or a microscope. In Fig. 15·1 we show the images and also the intensity contours of a *given pair* of point sources by *different* instruments. It will be seen that (*a*), (*b*) and (*d*) resolve the images (show separately), (*c*) and (*e*) do not resolve. It will be noted that separation of image centres is the same in (*a*), (*b*), (*c*) (meaning equal magnification), but

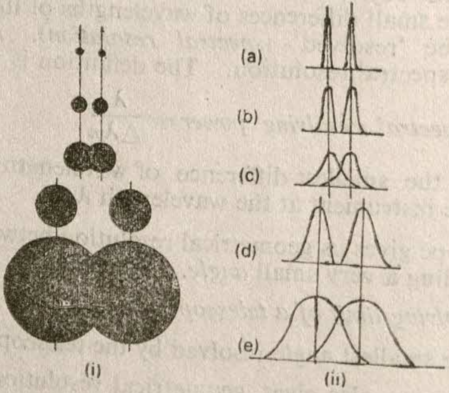


Fig. 15·1.
Cases of different magnifications and different resolutions

the intensity spreads are different, hence unequal resolution. Compare (*a*) and (*d*)—the latter has almost 3-fold magnification compared to the former. yet the resolution is much poorer. Compare (*c*) and (*d*)—the latter has higher magnification, and also higher resolution.

The intensity distribution curves of Fig. 15·1 also apply to images in a spectrum ; and thus spectral resolution or non-resolution is also expressed by the same curves. In Fig. 15·2 we show a given pair of spectral lines (say sodium D_1, D_2 lines) as seen through two different spectrographs. The *dispersion* (separation of centres) is more in (*a*) than in (*b*). But the intensity distribution curve shown just above (*a*) has only a small dip, whereas that shown above (*b*) has a larger dip. Thus the *resolution* is less in (*a*) than in (*b*).

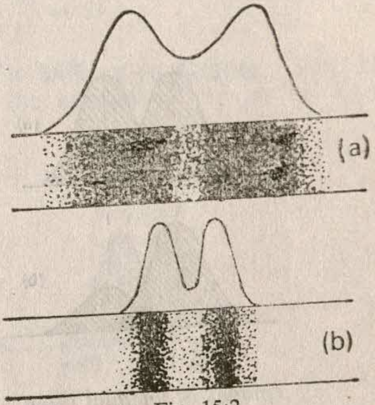


Fig. 15·2.
Spectrum in (*a*) *has more magnification but less resolution than that in* (*b*)

It may be noted that *if the images are unresolved, then no amount of magnification can ever show them resolved.* Therefore, the discussion of resolution centres on the primary image.

15·3. Rayleigh Criterion for limit of Resolution. Consider the three intensity distribution curves in Fig. 15·3. In all cases the images overlap. *i.e* , they are not *totally* separate. Yet in case (*a*) one can easily see that there are two images, case (*b*) is marginal, while in case (*c*) it is rather difficult. It is a personal judgment whether you call the images' resolved' or not.

To have a definite quantitative rule, Rayleigh laid down the following criterion for resolution, which is generally adopted : *Two images are just resolved if the central maximum of intensity curve for one image falls at the first minimum about the other image.* In Fig. 15·3 (*a*) the minimum of the left curve has not gone upto the central maximum of the right curve, but in case 15·3 (*c*) it extends even beyond the central maximum of the right curve. Hence by Rayleigh criterion images in (*a*) are resolved, while those in (*c*) are not resolved. Case (*b*) fits the Rayleigh limit.

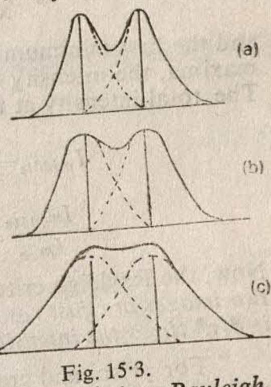


Fig. 15·3.
For debning Rayleigh limit of resolution

Rayleigh criterion applies only for two images of *equal intensity*. Consider cases (a) and (b) in Fig. 15·4. In both of them the separation of the centres of the images is equal to half-width of one of them. But while in (*a*) the intensities are shown equal, in (*b*) the intensities are very unequal. The resultant intensity cure (dotted) shows a dip in case (*a*), but no dip in case (*b*), so that in case (*b*) the weaker image gets lost in the tail of the stronger one.

Still, for comparing the resolving capacity of instruments, the Rayleigh criterion is taken as a standard.

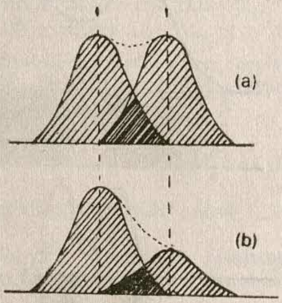


Fig. 15·4.

Images of unequal intensity and the Rayleigh criterion

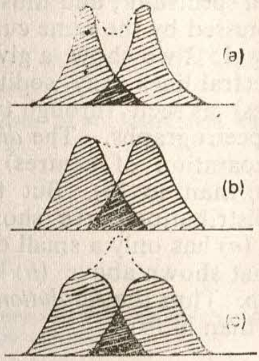


Fig. 15·5.

To show need of a modified Rayleigh criterion for resolution

However, the criterion has to be put in a modified form for another areson. Consider the intensity distributions of two images shown in fig. 15·5 at (*a*) (b) and (*c*). The separation of their centres and half-widths of images are the same in all cases (*a*), (*b*), (*c*) but shapes of the curves are different. As compared with case (*a*), the intensity distribution peak in case (*b*) is broad, while in case (*c*) it is flat. Looking at the resulant intensity curves (dotted), we find that compared with case (*a*), case (*b*) shows a far weaker dip, while in (*c*) we have no dip, but a hump. Hence the resolution is in fact best in (*a*) and worst in (*c*). For such reasons the Rayleigh criterion has to be put in a modified form.

$$I=I_0\frac{\sin^2 x,}{x^2}\ [I_{max}=I_0.]$$

and the first minimum occurs for $x=\pi$. Hence midway between the maxima, the intensity due to each source is given by putting $x=\pi/2$. The total intensity at the middle is therefore given by

$$I_{middle}=2I_0\frac{\sin^2\pi/2}{(\pi/2)^2}=\frac{8}{\pi^2}\,I_{max}$$

$$\frac{I_{middle}}{I_{max}}=\frac{8}{\pi^2}=0{\cdot}81 \text{ approx.} \qquad \ldots\ 15{\cdot}4)$$

Now, the Rayleigh criterion may be put in the modified from that *two images are just resolved if the intensity at the dip in the middle is $8/\pi^2$ times the intensity at either of the maxima.*

For prism and grating spectra, and also closely for telescope and microscope images, the two forms of the Rayleigh criterion are

Example 15·3. Imagine an intensity distribution of the form $I/I_0 = \sin^4 x/x^4$. Deduce $I_{middle} \div I_{max}$ when Rayleigh's original criterion is satisfied.

Solution. The first minimum occurs at $x=\pi$. Hence, at the 'middle', $x=\pi/2$, substitution gives

$$\frac{I_{middle}}{I_{mar}} = 2.\ \frac{\sin^4 \pi/2}{(\pi/2)^4} = \frac{32}{\pi^4} = 0.33.$$

15·4. Resolving Power of a Diffraction Grating. Fig. 15·6 shows the intensity contours of the spectral lines due to λ and $\lambda + \triangle\lambda$.

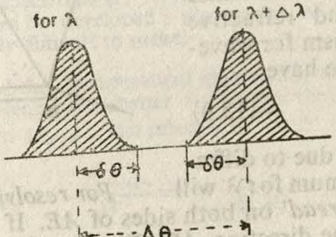


Fig. 15·6.

For discussing Resolving power of a spectral instrument

The angular separation of their peaks is $\triangle\theta$, and each line has a halfspread $\delta\theta$.

Now, for a grating the *separation* $\triangle\theta$ is given by eq. (14·20) as

$$\triangle\theta = \frac{n}{e \cos\theta} \triangle\lambda \qquad \ldots(15·5)$$

Also, if the grating has N lines, the spread $\delta\theta$ is given by Eq. (14·18) as

$$\delta\theta = \frac{\lambda}{N e \cos\theta} \qquad \ldots(15·6)$$

Combining eqs. (15·6) we obtain

$$\frac{\triangle\theta}{\delta\theta} = \frac{Nn}{\lambda} \triangle\lambda \qquad \ldots(15·7)$$

The *smallest* $\triangle\lambda$ which can be resolved is given by the Rayleigh condition $\triangle\theta_{min} = \delta\theta$. Putting this limit in eq. (15·7), we get

$$\text{Resolving Power} = \frac{\lambda}{\triangle\lambda_{min}} = Nn \qquad \ldots(15·8)$$

Thus, *the resolving power of a grating is equal to the product of the order of the spectrum and the total mumber of lines in the grating.*

The product *Ne* gives the total width of ruled space in the grating. Also if n_{max} is the maximum order of spectrum, we have $n_{max} \sim e/\lambda$. There fore, it is sometimes stated that the *maximum* resolving power of a given grating is ≈ total width÷λ.

15.5. Resolving Power of a Prism. In Fig 15.7, *ABC* is a prism and *BD* is the wavefront for the incident light. For light of wavefront λ the emergent wavefront is *CE*, which means that for this wavelength the optical path $DA+AE$ must be equal to optical path *BC*. If base length $BC=t$ and refractive index of the prism for wavelength λ is μ, we have

$$DA+AE=t.\ \mu \qquad \text{...(15.9)}$$

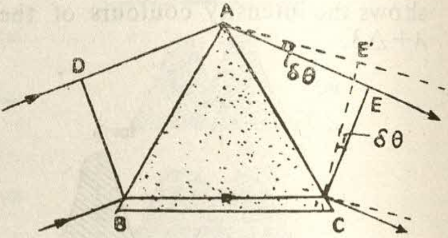


Fig. 15.7.
For resolving power of a prism

However, due to diffraction, the maximum for λ will have a finite '*spread*' on both sides of *AE*. If the minimum on one side occurs in the direction *AE'*, Fig. (15·7), the condition is that the extreme wavelets over plane *CE'* must have path difference λ, *i.e.*,

$$DA+AE'=\mu BC-\lambda=(t.\ \mu)-\lambda \qquad \text{...(15·10)}$$

For another wavelength $\lambda+\triangle\lambda$ the peak intensity may fall at any angle from *AE* depending on $\triangle\lambda$. But if $\triangle\lambda_m$ is the smallest difference of wavelength resolved, then, by Rayleigh criterion, the peak for $\lambda+\triangle\lambda_m$ falls just along *AE'*. The refractive index of the prism for this wavelengh is given by

$$\mu+\frac{d\mu}{d\lambda}\cdot\triangle\lambda_m.$$

Since its maximum lies along *AE'*, optical path $DA+AE'$=optical path *BC* for the wavelength $\lambda+\triangle\lambda_m$.

$$\therefore \qquad DA+AE'=t\left(\mu+\frac{d\mu}{d\lambda}\triangle\lambda_m\right) \qquad \text{...(15·11)}$$

From (15·10) and (15·11) we get

$$t\,\mu-\lambda=t\left(\mu+\frac{d\mu}{d\lambda}\triangle\lambda m\right)$$

$$\therefore \qquad \text{Resolving power}=\frac{\lambda}{\triangle\lambda_m}=-t\,\frac{d\mu}{d\lambda} \qquad \text{...(15·12)}$$

Thus the *resolving power of a prism is equal to the product of effective length of its base and the dispersive power* $|\,d\mu/d\lambda\,|$ *of its material.* For dense flint glass, used in spectrometers, the dispersive power is ~ 1000 cm^{-1}. With a large prism of 10 cm base the R.P. would be 10,000. At $\lambda=6000\ A$, the smallest $\triangle\lambda$ resolved would be 0·6 *A*.

Study of very small wavelength differences is of great theoretical interest. The two yellow lines in mercury spectrum (5770, 5790 *A*) require a resolving power of ~ 300 ; the two yellow in sodium (5890, 5896 *A*) require a resolving power of ~1000. But in some cases we need to resolve spectral lines as close as 0·02A at~5000A. That needs a R.P=250,000.

A typical grating spectrometer may have R.P. of about 50,000. In larger gratings one may go upto 200,000. For going still farther we need special devices like Fabry-Perot etalon, Lummer-Gehrcke plate etc. (§ 15·6)

In contrast, about the largest prism spectrometers have R.P. of 15,000 only (in the visible spectrum). Thus, as regards resolution, it is poorer than even the most ordinary grating spectrometer. It is unlikely that materials of $|d\mu/d\lambda|$ much higher than~1500 can be developed ; also prisms of bases longer than, say, 15 cm are exceptionally difficult to make.

It may be remarked that in theoretical physics, the wavenumber $\nu=1/\lambda$ is more important than λ. Hence it is better to express the resolution limit in terms of $\Delta\nu_{min}$ rather than $\Delta\lambda_{min}$. The relation is

$$|\triangle\nu_{min}|=\frac{\triangle\lambda_{min}}{\lambda^2}=\frac{1}{\lambda\times\text{R.P.}}=\frac{\nu}{\text{R.P.}}\qquad\ldots(15\cdot13)$$

$$\therefore\qquad \text{R.P.}=\frac{\lambda}{\triangle\lambda_{min}}=\frac{\nu}{\triangle\nu_{min}}.$$

It may be noted if λ and $\Delta\lambda$ are expressed in cm then ν and $\Delta\nu$ would be in cm^{-1}. (The name Kayser is given to the unit 1 cm^{-1}).

Example 15·4. A diffraction grating with 20,000 lines is used in the third order in range of wavelength 5500 *A*. What is the smallest $\Delta\lambda$ it can resolve? Calculate also the $\Delta\nu$ resolved.

Solution. $\text{R.P.}=Nn=20{,}000\times3=60{,}000$

$$\therefore\qquad \frac{5500\ A}{\Delta\lambda \text{ in } A}=60{,}000\Rightarrow\Delta\lambda=0\cdot091\ A$$

$$|\Delta\nu|=\frac{\Delta\lambda}{\lambda^2}=\frac{1}{\lambda\times\text{R.P.}}\Rightarrow\Delta\nu=0\cdot3\text{ cm}^{-1}.$$

Example 15·5. The slit of a spectrometer has width 2×10^{-3} mm, and the collimator has focal length 30 cm A 2 inch grating with 16000 lines per inch forms a spectrum with normal incidence. Deduce the contributions to the angular width of the second order spectral image for $\lambda=5000\ A$ due to (i) the slit width, (ii) the grating.

Solution. Angular width due to slit width$=\dfrac{2\times10^{-4}\text{ cm}}{30\text{ cm}}$

$$=0\cdot7\times10^{-5}\text{ rad.}$$

The second order spectrum falls at angle given by

$$\sin\theta=\frac{n\lambda}{e}=\frac{2\times5\times10^{-5}}{2\cdot54/16000}=\frac{1\cdot6}{2\cdot54}=\cdot63\Rightarrow\cos\theta=0\cdot78$$

From eq. (15·8), angular width due to diffraction is

$$\delta\theta = \frac{\lambda}{Ne\cos\theta} = \frac{5\times10^{-5}}{2\times2{\cdot}54\times{\cdot}78} = 1{\cdot}3\times10^{-5}\text{ radian.}$$

Thus, total width will be $2{\cdot}0\times10^{-5}$ radian. ⊙

Example 15·6. The refractive indices of extra dense flint glass at $\lambda=$ 6500, 6000, 5500 and 5000 *A* are 1·6420, 1·6464, 1·6516 and 1·6600. Deduce the resolving power of a prism of 8 cm effective base for $\lambda=6200$ *A* and for $\lambda=5200$ *A*. Also deduce the resolving limits $\Delta\nu$.

Solution. $d\mu/d\lambda$ may be obtained from a graph of μ versus λ. As a close approximation,

$$\left|\frac{d\mu}{d\lambda}\right| \text{ at } 6200\ A = \frac{1{\cdot}6464-1{\cdot}6420}{500\times10^{-8}} = 880\text{ cm}^{-1}$$

$$\left|\frac{d\mu}{d\lambda}\right| \text{ at } 5200\ A = \frac{1{\cdot}6600-1{\cdot}6516}{500\times10^{-8}} = 1680\text{ cm}^{-1}$$

Hence the resolving powers at the two given wavelengths are

$$8\times880 \quad\text{and}\quad 8\times1680.$$

$\Delta\nu$ is given by $1/(\lambda\times\text{R.P.})$. Simple substitution gives

$$\Delta\nu = 18\text{ cm}^{-1} \quad\text{and}\quad 12\text{ cm}^{-1}$$ ⊙

15·6. Some Devices for High Spectral Resolution. (*a*) *Fabry and Perot etalon.* In § 12·9 we described the *F.P.* interferometer, the intensity distribution in which is given by eq. (12·28), so that very sharp maxima occur as shown in Fig. 12·20. The *FP* etalon has the same description, except that both the 'mirrors' are at a fixed separation *e* (a few mm).

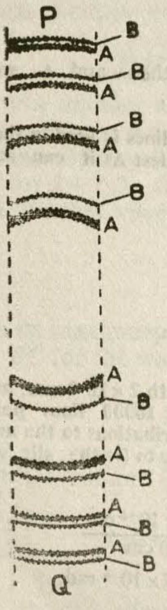


Fig. 15·8.
Fringes when a F-P etalon crosses a prism spectrum

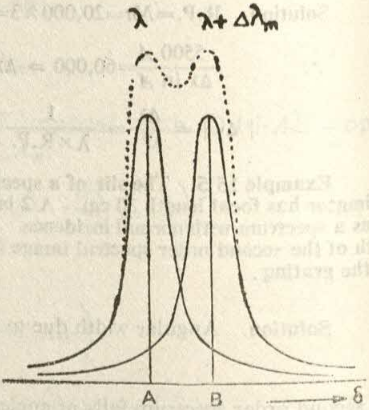


Fig. 15.9.
For discussing R.P. of a Fabry and Periot etalon

The *F.P.* etalon is used along with a prism spectrometer, being placed generally between the collimator and prism. In Fig. 15·8 in the absence of the *FP* etalon the strip *PQ* represents a spectral line as given by the prism. There is no evidence whether it contains a few lines too close to be resolved by the prism. When the *FP* etalon is inroduced the single vertical bright line *PQ* changes to a number offringes—apparently these are parts of concentric circles, and there is a double set, A and B, belonging to wavelengths λ and $\lambda+\triangle\lambda$. We have to determine $\triangle\lambda$.

Let $\triangle\lambda_0$ be such that n_{th} order for $\lambda+\triangle\lambda_0$ falls on the next $(n+1)_{th}$ order for λ. Then

$$n\,(\lambda+\triangle\lambda_0)=(n+1)\lambda \Rightarrow \triangle\lambda_0=\frac{\lambda}{n}=\frac{\lambda^2}{2e} \qquad \ldots(15{\cdot}14)$$

If now the separations of nearest *AB* and *AA* (Fig. 15·8) are measured, the $\triangle\lambda$ value is given by

$$\triangle\lambda=\triangle\lambda_0.\left(\frac{AB}{AA}\right)=\triangle\lambda_0 f \qquad \ldots(15{\cdot}15)$$

where $f=AB/AA$. The crucial feature is that since the maxima are very sharp, f as small as $\sim 1/20$ can be resolved. From eqs. (15·15) and (15·14), with f_m for the minimum f observable, we get

$$RP\equiv\frac{\lambda}{\triangle\lambda_m}=\frac{\lambda}{\triangle\lambda_0 f_m}=nN \qquad \ldots(15{\cdot}16)$$

where $\lambda/\triangle\lambda_0=n$ and we have put $1/f_m$ as the effective number of interfering beams N in the F.P. etalon.

Example 15·7. A Fabry-Perot etalon has plates of reflectivity 0·85 and separation 4·0 mm. Deduce the resolving power for $\lambda\approx 5000$ $A°$.

Solution. First, eq. (12·29) gives

$$F=\frac{4\times 0{\cdot}85}{(1-0{\cdot}85)^2}=150$$

Next, Fig. 15·9 with Rayleigh condition tells that at separation $\frac{1}{2}AB$,

$$\frac{I}{I_0}=\frac{1}{1+F\sin^2(\delta/2)}\approx 0{\cdot}405$$

Substitution for F leads to

$$\delta\approx 2{\cdot}5/\sqrt{F}\approx 0{\cdot}20 \text{ radian}$$

Since *AB* corresponds to 2δ and *AA* to 2π phase difference, we get

$$f_m=2\delta/2\pi\approx 1/16 \Rightarrow N=1/f_m=16.$$

$$n=2e/\lambda=12000 \quad \Rightarrow RP\approx 190{,}000$$

(*b*) *Lummer and Gehrcke Plate.* This is an optically plane plate of glass PQ (Fig. 15·10), with one end prismatic, such that the incident beam suffers total reflection at the prism face and then falls on the upper face of the plate at angle of incidence *r just short of* the critical angle. A typical one may have length $L=10$ cm, thickness $e=5$ mm and lateral width$\sim$1 cm.

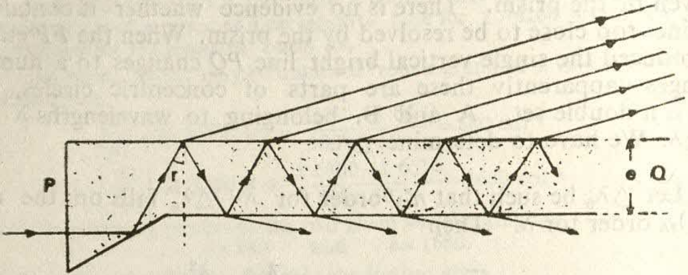


Fig. 15·10.

Lummer and Gehrcke Plate

For r near the critical angle the reflectivity becomes very close to unity. Therefore the several emergent beams on both sides (figure shows on one side only) may be treated as of nearly equal amplitudes. Since path difference between successive coherent beams is $2\mu e\cos r$ (eq. 12·5) the maxima are formed according to

$$2\mu e\cos r=n\lambda$$

The Lummer-Gehrcke plate (like the F.P. etalon) is used in conjunction with a prism spectrometer, so that a pattern like that of Fig. 15·8 appears.

For deducing the resolving power we shall not go into the intensity distribution, but use the rule-of-thumb formula R.P.$=nN$. For plate length L simple geometry of Fig. 15·10 gives $N=L/2e\tan r$, whence

$$\text{R.P.}=nN\approx\frac{2\mu e\cos r}{\lambda}\,\frac{L}{2e\tan r}\approx\frac{L(\mu^2-1)}{\lambda}\qquad\ldots(15\cdot17)$$

since r is nearly $\sin^{-1}(1/\mu)$. In the approximation taken, R.P. is independent of thickness e. But R. P. really equals nN_{eff} and this N_{eff} falls short of N by larger and larger factor as e decreases.

(*c*) *Michelson's Transmission Echelon*. Another device using large order of interferene (n) with small effective number of beams (N) to get high spectral resolving power is the Michelson's echelon. Fig. 15·11 shows the transmission type. A series of glass plates, each of thickness t, are arranged as a flight of steps of width e each. N is the number of steps, and n is in the neighbourhood of $(\mu-1)\,t/\lambda$. The theory is exactly that of a plane diffraction grating. But high n makes overlapping of orders a serious problem and the echelon has to be *crossed* with another low resolution device, as in the case of the F-P etalon or L-G plate.

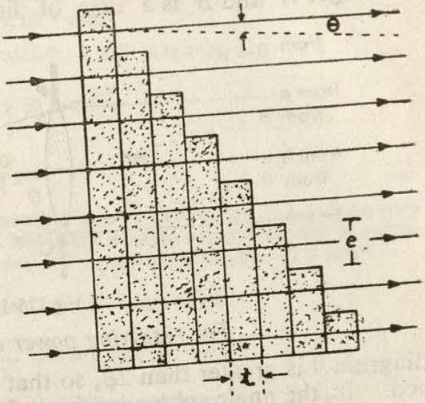


Fig. 15·11.
Michelson's Transmission Echelon

Reflection echelons have also been conceived in which the flight faces would be aluminised and the order of interference would be near $2t/\lambda$. These could be useful in regions where glass is not transparent. But in practice they have not been successfully made.

(*d*) **Fourier Transform Spectroscopy.** In Michelson's Interferometer the intensity distribution for an ideal single wavelength light (λ) is $\cos^2 k\delta$ type. If the line has finite spectral width (all practical sources *must* have), then the intensity distribution *correspondingly* changes. If there are two spectral lines, one gets consonance and dissonance [§12·6 (*b*)]. If there are many lines, then again the intensity distribution represents all that information—the spacing and relative intensities of the member lines.

Now the reverse process is Fourier transform spectroscopy (FTS). In Chapter 9 we saw that while. $f(t)$ is the integration of the components $F(k)$ multiplied with $\exp(ikt)\,dk$, we may go the opposite way to deduce $F(k)$ from the observed $f(t)$ by integration of $f(t)\exp.(ikt)\,dt$. In FTS the observed intensity distribution is $f(t)$; electronic circuits do the Fourier integration for different k values turn by turn, and the resultant $F(k)$ values are plotted against k. That is the spectrum. To eliminate errors, the intensity in the Michelson fringes is scanned across several different diameters and the average is subjected to Fourier transform process. It may be noted that FTS can be used only when the fringe profile of an ideal (zero-width) single line has a harmonic variation ($\cos k\delta$ type). Twin-beam interference alone satisfies this requirement, Michelson interferometer being one such case.

15·7. **Resolving Limit of a Telescope.** Resoving *limit* of a telescope is defined as *the smallest angle θ_{min} subtended at the observer's place by two distant objects* (*of equal brightness*) *which can be just seen as separate ones through the telescope*. The resolving *power* may be expressed by the reciprocal of θ_{min}.

In Fig. 15·12, lights coming from two distant objects A and B subtend an angle θ at the telescope objective PQ. The centres of their images formed by the objective are A', B' and their *angular* separation is $A'OB' = AOB = \theta$. Now, due to diffraction effect each of images A' and B' is a *disc* of finite angular radius ϕ. In this diagram θ is greater than 2ϕ, so that the disc images are well resolved. If the angle subtended by A, B (*i.e.*, angle θ) is smaller, the disc images would overlap. Rayleigh criterion requires that at the resolving limit,

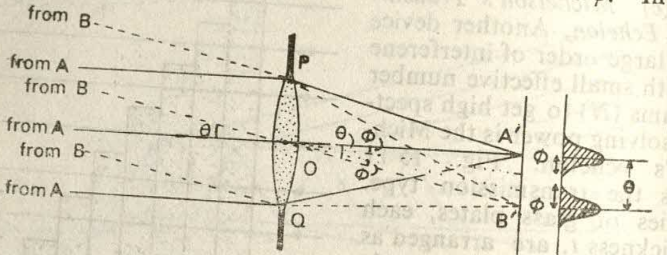


Fig. 15·12.

For resolving power of a telescope

$$\theta_{min} = \phi$$

But Fraunhofer diffraction at a circular aperture (§ 14·3) gives angular spread $\phi = 1{\cdot}22\lambda/d$, where d is *diameter* of the telescope objective and λ is the effective wavelength for observation. Hence we get,

$$\text{resolving limit} = \theta_{min} = 1{\cdot}22\,\frac{\lambda}{d}. \qquad \ldots(15{\cdot}18)$$

Note that the spread $\phi = 1{\cdot}22\lambda/d$ occurs even under the best optical conditions of image formation. Aberrations (if any) will cause further spread. Hence, actual resolution limit may be larger than that given by eq. 15·18.

In actual viewing through a telescope there is another significant factor for consideration. Let the resolving limit of the objective of the telescope be θ_T and that of the human eye be θ_E. Then although the objective resolves (in the primary image) two objects subtending the smallest angle θ_T, the eye will not be able to resolve them unless the manification M is such that

$$\theta_T . M \geqslant \theta_E$$

The least value of M desired, often called *normal magnification* M_0, is given by

$$\theta_T M_0 = \theta_E.$$

One can see that

$$M_0 = \frac{\text{aperture of the objective}}{\text{aperture of the eye}} \qquad \ldots(15{\cdot}19)$$

Example 15·8. Calcutate the smallest angular separation resolved by the human eye, given : aperture=2·5 mm, and effective $\lambda = 5500\ A$. If a scale with millimeter markings is viewed by the unaided eye, deduce the largest distance to which the markings will be visible.

Solution. $\theta_{min}=\dfrac{1{\cdot}22\lambda}{d}=\dfrac{1{\cdot}22\times5{\cdot}5\times10^{-5}\text{ cm}}{2{\cdot}5\times10^{-1}\text{ cm}}$

$$=2{\cdot}7\times10^{-4}\text{ rad.}$$

From distance x metre the angle subtended by 1 mm markings is

$$\theta=\frac{1\times10^{-1}\text{ cm}}{x\times\ 100\text{ cm}}=\frac{1}{x}\times10^{-3}\text{ rad.}$$

Equating this with θ_{min} we get

$$x\approx4\text{ metres.}$$ ⊙

Example 15·9. Calculate the smallest angle resolved by a telescope with an objective of 5·0 m diameter. What minimum magnification is needed to enable the human eye to utilise such high resolution ? (aperture of the eye=2·5 mm).

Solution. We take effective λ as $5{\cdot}5\times10^{-5}$ cm. Then

$$\theta_{min}=\frac{1{\cdot}22\times5{\cdot}5\times10^{-5}\text{ cm}}{500\text{ cm}}=1{\cdot}34\times10^{-7}\text{ rad.}$$

The minimum magnification M_0 needed is given by eq. (15·19) as

$$M_0=\frac{5{\cdot}0\times100\text{ cm}}{2{\cdot}5\times10^{-1}\text{ cm}}=2000$$ ⊙

Example 15·10.* A radar antenna is 3 metres across and the wavelength used is 1 cm. Estimate the limit of (i) angular resolution, (ii) linear resolution at 2000 metre distance.

Solution. $\theta_{min}=\dfrac{\lambda}{d}=\dfrac{1\text{ cm}}{300\text{ cm}}=\dfrac{1}{300}\text{ radian}$

$$x_{min}\sim2D\times\theta_{min}\sim13\text{ metre.}$$

[Factor 1·22 is omitted because only the order of magnitude is desired. Actual factor depends upon the shape of the radar antenna] ⊙

Example 15·11.† Ultrasonic waves of frequency 50,000 Hz are produced in water, the radiating disc having a diameter 40 cm. Calculate the width of the main beam at a distance of 100 metres. (Speed of sound in water=1500 m/sec.)

Solution. $\lambda=\dfrac{\text{speed}}{\text{frequency}}=\dfrac{1500\text{ ms}^{-1}}{50000\text{ s}^{-1}}=3\times10^{-2}m,$

$$\text{Angular spread}=1{\cdot}22\frac{\lambda}{d}=1{\cdot}22\times\frac{3\times10^{-2}\text{ m}}{40\times10^{-2}\text{ m}}=0{\cdot}92\text{ radian.}$$

Linear spread at 100 metres taking both sides of the normal ⊙

$$=2\times100\times{\cdot}092\ m=18\text{ m.}$$

15·8. Michelson's Stellar Interferometer.

Figure 15·13 shows the principle of a Michelson's stellar interferometer. The telescope objective is covered with two slits S_1, S_2, so that in the focal plane AB each point source shows interference fringes (not an image). Light enters through a system of mirrors M_1 to M_4. If there are two point sources separated by θ, then the two sets of interference fringes are in first dissonance when $D\theta=\lambda/2$ One varies D till the fringes disappear. That leads to $\theta=\lambda/2D$.

---

* This gives the principle cf radio detection of targets, called *RADAR* Without using waves of very short λ the beam would spread too much to be of any use.

† This gives the principle of SONAR (sound wave detection and ranging), used to locate objects under sea, like icebergs, submarines, etc. Again, very high frequency (small λ) is necessary to produce a beam of smaller angular spread.

Actually Michelson used the device not to measure the separation of two stars, but to determine the angular diameters of stars. (The expression is stightly different). Intensity loss is a serious problem; but to get high D values without making the lens aperture large is a tremendous advantage.

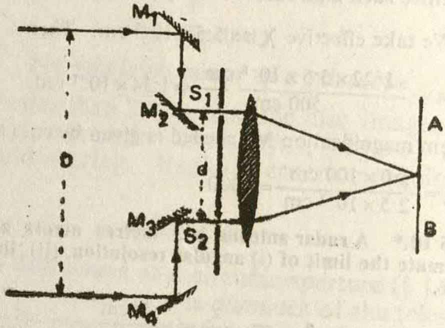


Fig. 15·13

The Stellar Interferometer

15·9. Resolution in a Microscope. If d_E is diameter of the entrance pupil of the human eye, the smallest angle between objects which it can resolve is $1{\cdot}22\lambda/d_E$. Further, the nearest distance from which the eye can distinctly see is $D=25$ cm. Hence, the smallest *linear detail* which an unaided eye can resolve is $x_E=1{\cdot}22\lambda D/d_E$. This comes to $\sim 0{\cdot}1$ mm.

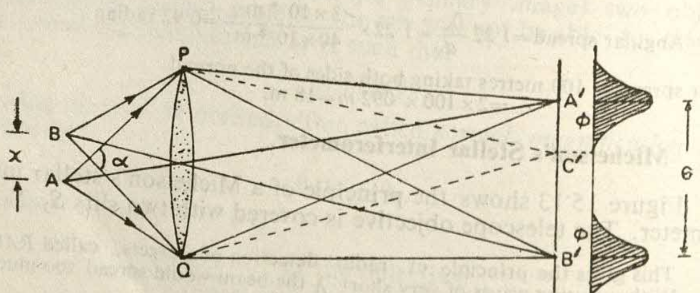


Fig. 15·14

Resolution in a Microscope (incoherent illumination)

In Fig. 15·14 the microscope objective is (for convenience) shown by a single lens PQ. Let A and B be two close object points with separation $AB=x$. The *centres* of their images are A' and B' respectively, and each image shows an angular spread ϕ as shown by the intensity curves on the right end. The figure shows A and B well resolved. As B goes nearer to A (x decreases), image B' goes nearer to image A'. In the Rayleigh limit of resolution B' would coincide with C', the minimum near A'. The corresponding $x=x_{min}$ gives the resolving limit of the microscope objective.

That C' is the first minimum around A' leads to relation

$$AP+PC'=AQ+QC'+1{\cdot}22\lambda \qquad \text{...(15·20 a)}$$

and that in the resolution limit B' would coincide with C' gives

$$BP+PC'=BQ+QC' \qquad \text{...(15·20 b)}$$

Combining these equations gives

$$(AP-BP)+(BQ-AQ)=1{\cdot}22\lambda \qquad \text{...(15·21)}$$

The geometry of fig. 15·13 shows that if α is the *semi-angle* subtended by the objective PQ at the object, we have $AP-BP=x_{min}\sin\alpha$ $=BQ-AQ$. Hence eq. (15·21) reduces to

$$x_{min}=0{\cdot}61\frac{\lambda}{\sin\alpha} \qquad \text{...(15·22)}$$

If the space between the object and the lens is filled with a liquid of refractive index μ, the effective wavelength becomes λ/μ, so that

$$\text{Resolving limit } x_{min}={\cdot}61\frac{\lambda}{\mu\sin\alpha} \qquad \text{...(15·23)}$$

The quantity $\mu\sin\alpha$ is called *Numerical aperture* of the objective. For an oil-immersion objective, it has been possible to reach numerical aperture about 1·6. With $\lambda\sim5500A°$, the resolution limit comes to

$$x_{min}\sim\frac{0{\cdot}61\times5{\cdot}5\times10^{-5}\text{ cm}}{1{\cdot}6}\approx2\times10^{-5}\text{ cm.}\ \approx2000\ \mathring{A}$$

It is worth noting that when all the chromatic or other aberrations in image formation are properly taken care of, there is still a fundamental limitation due to which each image *must* have some spread, which limits the resolution.

If objects of separation x_{min} are ultimately to be seen by the eye through the microscope, the magnification M should be such that

$$M\,x_{min}>x_E=\frac{1{\cdot}22\,\lambda D}{d_E}$$

The minimum magnification required, called *'normal magnification'* M_0, is given by

$$M_0=\frac{1{\cdot}22\,\lambda D}{d_E\,x_{min}}=\frac{2D}{d_E}\mu\sin\alpha \qquad \text{...(15·24)}$$

Electron Microscope. Examining eq. (15·23), we notice that for lowering the resolving limit there is little hope of improving 'numerical aperture' by any large factor. Hence, the only choice is to decrease λ. If ultraviolet light (*e.g.*, mercury 2536 A) is used in place of visible light, we may improve resolving limit by a factor of about 2. Modern theory [see chapter 22] tells us that electrons of momentum p behave as waves of wavelength λ given by

$$\lambda = \frac{h}{p} \qquad \text{(de Broglie equation)} \qquad \ldots(15\cdot25)$$

where h is Planck's constant. Typically, for electrons accelerated across a potential difference of 100 volt, $\lambda = 1\cdot2\ A$. It has been posssible to construct 'lenses' for focussing beams of electrons[*], and hence to construct the 'electron microscope'. Due to design problems its numerical aperture is small. Even then because of great reduction in λ the resolving limit is considerably improved. For a microscope using 100-volt electrons[⁑] and numerical aperture 0·1.

$$x_{min} = 0\cdot6 \times \frac{1\cdot2A}{0\cdot1} \sim 7A.$$

Electron microscepes are now widely used to examine the smallest details in materials like liquid films, coatings, bacteria, macromolecules, etc. They give information only about thin layers or surface coatings.

Ion microscopes. These work on the same principle as the electron microscope. They use a beam of ions. Since $p = \sqrt{2mE}$, you get a smaller λ for the same energy. That is the advantage. But lenses are more difficult to make.

Example 15·12. A microscope objective has aperture diameter 8 mm and focal length 2·1 cm. Estimate its resolving power.

Solution. We assume that effective $\lambda = 5\cdot5 \times 10^{-5}$ cm and also that the object has to be placed only a little beyond the focal distance ; so this distance may be taken as $\sim 2\cdot5$ cm. Then

$$\tan\alpha = \frac{0\cdot4}{2\cdot5} = 0\cdot16, \qquad \sin\alpha = 0\cdot16 \text{ (check it)}$$

$$\therefore\quad x_{min} \approx 0.61 \frac{\lambda}{\mu \sin\alpha} = 0\cdot61 \times \frac{5\cdot5 \times 10^{-5}}{1 \times 0\cdot16} \approx 2 \times 10^{-4} \text{ cm.}$$ ⊙

Example 15·13. A microscope has an oil-immersion objective, with a hemispherical lens of diameter 6 mm and $\mu = 1\cdot60$. Deduce the numerical aperture.

Solution. For an oil-immersion objective, the object distance is $1/\mu$ times radius (see fig. 6·12)

$$\therefore\quad \tan\alpha = \frac{0\cdot3}{0\cdot3/1\cdot6} = 1\cdot6. \qquad \sin\alpha = 0\cdot85 \text{ (check it)}$$

$$NA = \mu \sin\alpha = 1\cdot6 \times 0\cdot85 = 1\cdot36$$ ⊙

15·10. Resolution with Coherent Illumination : Abbe's Theory

The theory of § 15·8 applies where the objects are self-luminous, or are illuminated by a source sharply focussed on them. Then waves from separate points of the object are not coherent and *intensity contours* for different parts are added. However,

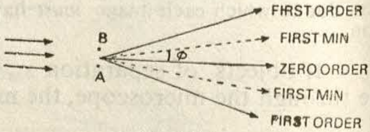


Fig.15.15 (*a*)
For Abbe's theory

---

[*] The 'electron lenses' are actually current-carrying coils, whose magnetic fields deflect the moving electrons.

[⁑] Larger voltage would mean smaller λ. But 'lenses' then become more difficult ; also the beam then damages the sample itself.

in most cases, the object is illuminated diffusely, so that waves from different neighbouring objects have definite phase relationships (coherent). For such cases Abbe has given a different theory. He considers a pair of close points, such as A and B (fig. 15·15), as a diffracting element of width $AB=x$. The first minima on either side of the zero order maximum occur at angle ϕ given by

$$x \sin \phi=\lambda.$$

Abbe now states that *the separate existence of A and B is realised through the microscope if at least the first minimum on either side of the central maximum participates in image formation.*

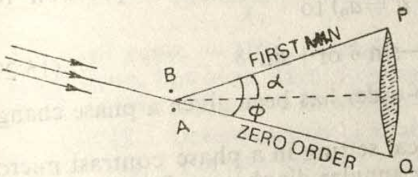


Fig. 15·16
increased resolution by oblique illumination

This means that (for axial illumination) in the limit the semi-angle α subtended by the microscope objective at A must be given by $\alpha=\phi$

$$\therefore\ x_{min} \sin \alpha=\lambda$$

or $$x_{min}=\frac{\lambda}{\sin \alpha}$$

For an oil immersion objective effective wavelength is λ/μ.

$$\therefore \quad x_{min}=\frac{\lambda}{\mu \sin \alpha} \qquad \ldots(15{\cdot}26)$$

In fig (15·16) the illumination is not along the axis, but along the edge AQ. In this situation the 'first minimum' is able to enter the objective upto ths limit $\phi=2\alpha$. This limiting substitution, including possible oil immersion, gives ue

$$x_{min}=\frac{\lambda}{\mu \sin 2\alpha} \qquad \ldots(15{\cdot}27)$$

It may be noted that eqs. (15·23), (15·26) and (15·27) all tell basically the same story, except for a numerical factor which depends on the manner of illumination. In fact, proper illumination of the object is itself a complicated question in microscopic vision.

15·11. Phase Contrast Microscope. Frequently we come across specimen which do not show variation in intensity in the transmitted light, but variations in phase—whether due to differences in thickness or of refractive index. Images in such cases show only the outlines, not the details. In biological specimen we have the regular art of 'staining' the sample by suitable dyes, which gives different tinges of colour to different parts of the specimen in order to identify them.

Zernike developed a microscope in which phase variation ,in passing through different parts of the specimen are converted into intensity variations in the image. This is called the *phase contrast microscope*. The principle is as follows. The centre of the image of any part has contribution from the zero-order diffraction due to that part, superposed by the first order diffraction from the neighbouring parts. Under *coherent illumination* the resultant disturbance is obtained as

$$a_0 \cos \omega t+a' (\cos \omega t-\delta)$$

where a_0 is the amplitude due to the zero-order and a' that due to the first order. Here δ is the phase difference to be converted into intensity variation.

By the interference theory (see eq. 11·5 *a*) *fractional* intensity change f for phase difference δ is given by

$$f=\frac{\Delta I}{I}=\frac{2a_0\, a'\cos\delta}{a_0^2+a'^2} \qquad \ldots(15{\cdot}28)$$

We note that f will be large if $a'=a_0$. That is one aspect. We further see that $df/d\delta$ comes out proportional to $\sin\delta$, which tends to zero for small δ. To get large $df/d\delta$ we would be advised to cause an extra phase change $\pi/2$ in the zero-order beam. That is the second aspect. These two steps together modify the expression for fractional intensity change (for $a'=a_0$) to

$$\frac{\Delta I}{I}\,\alpha-\sin\delta \text{ or} +\sin\delta \qquad \ldots(15{\cdot}29)$$

depending on whether the zero-order has been given a phase change $+\pi/2$ or $-\pi/2$.

Fig. 15·17 shows the typical setting in a phase contrast microscope. Light through a narrow annular diaphragm D is focussed on the specimen, and a matching *phase plate P* is introduced at the

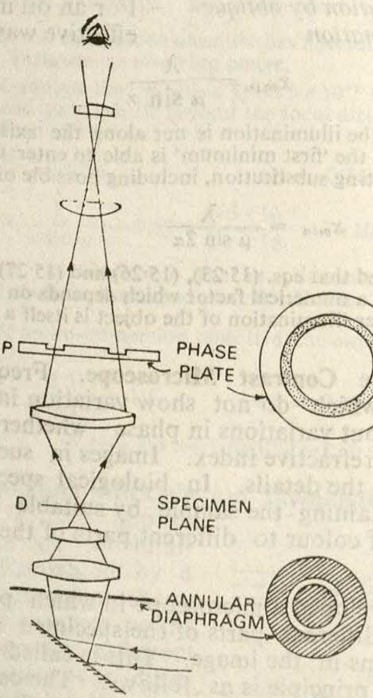


Fig. 15·17.
Phase Contrast Microscope

position where the image of D is formed by the intervening lenses. The phase plate has a ring-type groove or elevation to cause an extra phase difference $-\pi/2$ or $+\pi/2$ relative to rest of the plate. Also, this groove or elevation has a thin metallic deposit to reduce a_0 to make it match a'.

Since the phase plate *P* is in a position conjugate to the annular diapragm *D*, the zero-order light passes through the groove or elevation in *P* and light of other orders in the diffiraction passes through rest of the plate. Eq. (15·29) means that for small phase changes δ we have $\triangle I/I$ proportional to δ. Typically for a specimen 0·01 mm thick a variation 0·001 in refractive index would amount to

$$\delta = \frac{2\pi}{\lambda}\triangle\mu\, t \approx \frac{6}{6\times10^{-5}}\times 0{\cdot}001\times 0{\cdot}001 \text{ cm} \approx 0{\cdot}1 \text{ rad}$$

This will cause $\sim$ 10% intensity change. Such is the sensitivity. It follows, however, that if changes in *t* or *n* of the specimen are large, serious difficulties in interpretation can also arise.

15·12. Determination of Spectral Resolving Power. (*a*) **Diffraction grating.** For this purpose, typically, the spectrum of a sodium lamp is obtained in the desired 'order' with the given grating. It shows the two yellow lines $\lambda_1 = 5895{\cdot}93$ *A*, $\lambda_2 = 5889{\cdot}97$ *A* for which $\lambda_{mean} = 5893$ *A* and $\triangle\lambda = 5{\cdot}96$ *A*. Now a variable graduated slit is mounted on the face of the *collimator*. This slit is gradually narrowed down, till we find that the two yellow lines in the spectrum merge into one. The width of the slit for this setting is measured. Let this be *w*. Let *W* be the total width of the *ruled part* of the grating. Then we have

$$\text{Resolving Power} = \frac{5893}{5{\cdot}96}\times\frac{W}{w} \qquad \ldots(15{\cdot}30)$$

The grating must be used with normal incidence. If the variable slit is used on the telescope objective, *w* must be replaced by $w/\cos\theta$, where θ is the angle of diffraction.

(*b*) **Prism.** The same arrangement may be used for a prism spectrometer, except that a mercury source is more desirable. It has

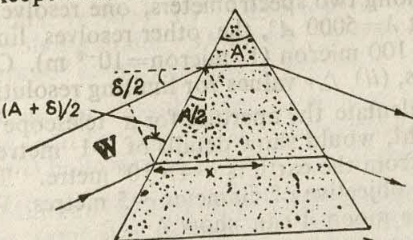


Fig. 15·18.

Determining R. P. of a Prism

a yellow doublet (5790·65, 5769·60 *A*), with $\lambda_{mean} = 5780$ *A*° and $\triangle\lambda = 21{\cdot}0$ *A*. If these lines merge when the beam width is reduced to *w*, the R.P. is calculated in two steps, Fig. 15·18 shows that the effective prism base *x* corresponding to beam width *w* is given by

$$x=\frac{w}{\cos\frac{1}{2}(A+\delta)}\,(\sin\tfrac{1}{2}A)\times 2$$

For the full prism of base t we then have

$$\text{Resolving Power}=\frac{5780}{21{\cdot}0}\times\frac{t}{x} \qquad \cdots(15{\cdot}31)$$

15·13. Determination of Resolving Power of a Telescope. Several black stripes of equal width a and equal spacing b are drawn on a white paper. This design is placed a few metres away from the telescope, with its plane perpendicular to the axis of the telescope. The design is focussed in the telescope. Now, a variable graduated slit is mounted on the telescope objective, with its length parallel to the stripes. The slit is narrowed gradually till the design shows uniform shade, not black and white stripes. Let this critical width of the slit be w. Let d be the total diameter of the telescope objective, and D the distance from the objective of the telescope to the drawing. Then,

$$\text{For aperture } w\text{, smallest angle resolved}=\frac{a+b}{D}$$

$$\therefore \quad ,, \quad d, \quad ,, \quad ,, \quad \theta_{min}=\frac{a+b}{D}\times\frac{w}{d}\times 1{\cdot}22$$

The factor 1·22 is introduced because the aperture d is circular, while w is for a rectangular slit.

It may be noted that $a+b$ appears in the calculation, not just the white spacing a or the black width b. It is not necessary to have $a=b$, although the judgement is difficult when a and b are very much unequal.

It is instructive to do the same experiment with a distant wire-netting if available. If a square netting with, say, 5 wires per cm is fitted on a distant window or door, say, 20 metres away then an ordinary telescope will resolve it with full aperture, but not when the aperture is reduced below about 6 to 8 mm.

PROBLEMS

15·1. Among two spectrometers, one resolves lines of separation 0·20 $A°$ at $\lambda=5000\ A°$, the other resolves lines of separation 0·02 micron at 100 micron (1 micron$=10^{-6}$ m). Compare their (*i*) resolving powers, (*ii*) $\triangle\nu$ values for limiting resolution.

15·2. Calculate the aperture for a telescope objective which, using visible light, would show details of $\sim$1 metre on the moon, whose distance from the earth is $\sim 4\times10^{8}$ metre. The largest telescope in use has objective of diameter $\sim$5 metres. What is the smallest detail on the moon it can show ?

15·3. Two particles of diameters 1.5×10^{-4} cm and $0{\cdot}5\times10^{-4}$ cm are seen through a microscope of resolving limit $1{\cdot}2\times10^{-4}$ cm and magnification $40x$. Deduce the angular diameters of their images, given least distance of distinct vision as 25 cm.

15·4. A radar antenna lying horizontal has length 20 m along north-south and 5·0 m along east-west. It uses waves of $\lambda=2{\cdot}0$ cm. Calculate its resolving limit in the north-south and east-west vertical nes.

15·5. A diffraction grating has 5000 lines per cm, and the total ruled width is 5 cm. Calculate for $\lambda=6000\ A$ in second order (*a*) the dispersion, (*b*) the resolving power, (*d*) the smallest $\triangle\lambda$ resolved, and (*d*) the smallest $\triangle\nu$ resolved.

15·6. A diffraction grating used with normal incidence gives the first order spectrum of a spectral line at 48°. What happens if (*i*) half the ruled width is covered from one end, (*ii*) two thirds of the number of rulings in the middle part of the grating are covered ?

15·7. Light of $\lambda=4{\cdot}31\times10^{-5}$ cm is incident normal to a diffraction grating of $e=1{\cdot}62\times10^{-4}$ cm and $N=60000$. For the third order spectrum deduce (*i*) the angular position of the spectral line, (*ii*) the dispersion $\triangle\theta/\triangle\lambda$, (*iii*) the angular half-spread of the line, and (*iv*) the smallest $\triangle\lambda$ resolvable.

15·8. A concave grating with 4000 lines per cm, total ruled with 15 cm, and radius of curvature 2 m is used in Rowland mounting. Calculate (*a*) linear dispersion in second order for $\lambda=5000\ A$, (*b*) the resolving limits $\triangle\lambda$ and $\Delta\nu$. How would these values be affected if we change in turns (*i*) the number of lines per cm, (*ii*) total ruled area, (*iii*) the radius of curvature ?

15·9. A plane grating just resolves the sodium doublet, $\Delta\lambda=6A$, when the effective width is reduced to 0·3 cm. What is the smallest $\Delta\lambda$ resolved at $4000A$ in the same order by the full grating in this region if total ruled width in 3·0 cm ?

15·10. A prism has base 4·0 cm and $d\mu/d\lambda$ of its material is $-860\ \text{cm}^{-1}$ at $\lambda=5460A$. Deduce the smallest $\Delta\lambda$ it can resolve in this region. Discuss how this value may change as we go towards the shorter wavelengths.

15·11. The Mount Palomar telescope has diameter 5 metres. The images of stars formed by it do not show sizes noticeably greater than the resolving limits. Deduce a conclusion about the angular sizes of stars (of course, excluding the Sun).

15·12. A microscope objective has numerical aperture 0.12. Deduce its resolving limit for (*i*) coherent illumination, (*ii*) incoherent illumination, (*iii*) hollow-cone illumination. ($\lambda=6\times10^{-5}$ cm).

15·13. In a Millikan oil-drop experiment almost all the drops appear to be of the same size, though they fall under gravity with very different speeds. Interpret this in the context of resolving limit of the observing microscope.

15·14. A person claims that he saw a man 50 m away wearing a shirt with stripes about 1 cm wide. Comment on this.

15·15. A wire gauge, with its plane vertical, has horizontal wires at 2 mm interval and vertical wires at 4 mm interval. It is seen in daylight from 100 m distance through a telescope of aperture diameter 2·5 cm. Discuss what details of the gauge are seen. What would happen if a slit of width 1 cm is placed in front of the telescope with its length (*i*) vertical, (*ii*) horizontal ?

15·16. Compute the resolving power of a Fabry-Perot etalon of $F=64$ and spacing 6·0 mm near $\lambda=5000\ A$. Why can we not increase the *RP* by making the spacing (and hence *n*) very large ?

15·17. If the intensity distribution in a spectral image is $I/I_0=\sin^4 x/x^4$ (see Example 15·3), deduce the smallest phase angle *x* for which two images of equal intensity would be resolved according to Rayleigh's revised criterion.

15·18. An electron microscope using 100*V* electrons and numerical aperture 0·08 has resolving limit ~10A. Compute the resolving limit for a 400 *V* electron microscope of N.A.=0·02.

15·19. What is a *phase plate* of a phase contrast microscope, and where is it located ? Explain its function.

16

Polarization of Light and Double Refraction

The phenomena of interference and diffraction would occur with light whether the waves are transverse or longitudinal. But in Chapter 9 we have seen that waves of light are identical with electromagnetic waves, which are of transverse character. In this chapter we will deal with some special results arising from the transverse character of light.

16·1. Evidence that Waves of Light are Transverse. For electromagnetic waves in the microwave region ($\lambda \approx 1$ cm) the transverse character can be demonstrated using a *wire-grid polarizer.* It has a number of straight parallel wires (Fig. 16·1) of an electric conductor. Let the plane of the wires be xy plane with wires along x-direction, and let the waves travel along the z-direction. If one more wire-grid (not shown) is placed parallel to the first and is rotated in its own plane, then no waves pass if the wires in the two grids are mutually perpendicular, while waves pass if the wires in the two grids are parallel.

The explanation is simple. The x-oscillations of electric field E in the e.m. waves make the electrons in the wires oscillate and the energy is thus dissipated as heat. The y-oscillations of E field are unable to do this (assuming wire diameter $\ll \lambda$) and hence pass through.*

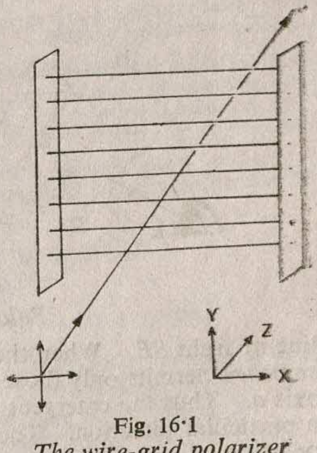


Fig. 16·1

The wire-grid polarizer

For visible light the transverse character may be established by an experiment like that described in fig. 16·2. Light from a source S is passed through two plates A and B of a crystalline material called *tourmaline.*† The receiver may be the eye or a photoelectric cell. If plate A is kept fixed and B is rotated in its own plane, then in each rotation there are two positions (at 180° separation) for which the transmitted light is large, and two other positions (at 90° from the previous positions) for which the transmitted light is almost zero. The same is observed if plate B is kept fixed and A is rotated in its own plane.

---

* With visible light ($\lambda \approx 5 \times 10^{-5}$ cm) the wire-grid polarizer would not work because we would need more than $\sim 10^4$ wires per centimeter width. But in infrared regions wire-grid polarizers are used.

† The same may be observed with a pair of polaroids or Nicols which we will later describe.

To understand the observations described above, imagine that source S gives vibrations as shown at P (lower diagram in fig. 16·2), that is, equally possible in all directions in a plane perpendicular to line of sight SE. When these pass through tourmaline plate A, it *somehow* permits only those vibrations which are along a particular axis a. Thus the emergent vibrations, shown at Q, are only along a particular direction. The next tourmaline plate B, in the given position, allows vibrations along an axis b only. If this axis is parallel to $a(\theta=0)$, the vibrations shown at Q pass through plate B fully. If axis b is at 90° to a, the vibrations shown at Q are totally cut off by B. This explains the observations.

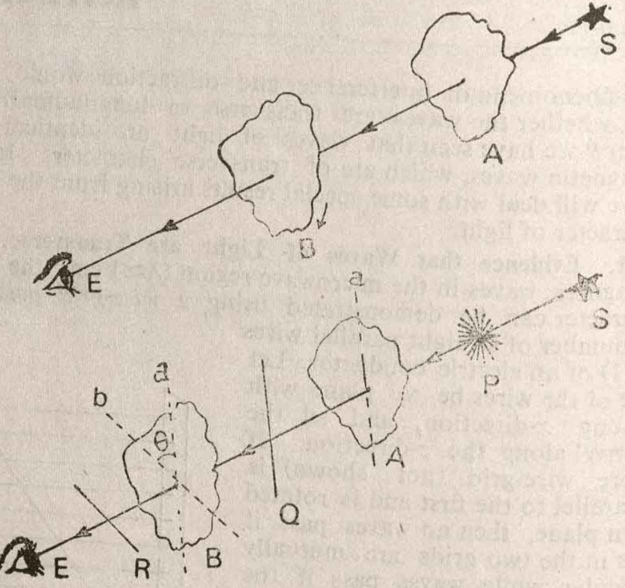


Fig. 16·2.

Polarization of light

However, this model makes a detailed prediction. If the angle between axes b and a is in general θ, then a fraction $\cos\theta$ in amplitude would be allowed through by plate B. Since intensity $\propto$ (amplitude)^2, we should have intensity varying as

$$I_\theta = I_0 \cos^2 \theta$$

This is called the *law of Malus*. It has been extensively verified, so that the model is satisfactory.

Fig 16·3 shows a simple arrangement. A collimated beam is passed through a pair of nicols N_1, N_2 (which we will later describe), one of which may rotate about the beam axis. P is a photo-cell measuring the intensity of the transmitted light as θ varies.

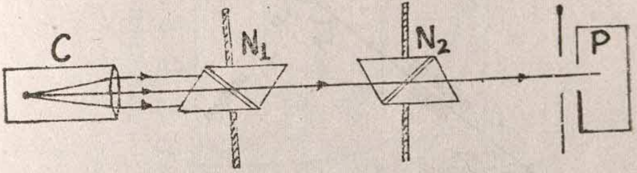


Fig. 16·3.

Experiment to check the law of Malus

The beam with vibrations shown at P in fig. 16·2 is said to be *unpolarized*, meaning that there is no preferred direction about the beam axis. The vibrations are randomly oriented with equal probability. The beam shown at Q is said to be *plane polarized*, meaning that the vibrations are only along normal to a particular plane* containing the beam axis. In between, we will later come across *partially polarized* light, containing vibrations unequally distributed in the plane normal to the beam.

A tourmaline plate is called a *polarizer*, so is a polaroid or a Nicol prism. In Fig. 16·3 the second Nicol (though similar to the first) is called an *analyzer*.

Example 16.1. An analyzer examines two adjacent plane polarized beams *A*, *B* whose planes of polarization are mutually perpendicular. In one position of the analyzer, beam *B* shows zero intensity. From this position a rotation of 27° shows the two beams as matched. Deduce the intensity ratio I_A/I_B of the beams.

Solution. In the zero position, as described, the vibrations transmitted by the analyzer are parallel to the vibrations in beam *A*. Hence, in the position 27° from this,

$$I_A \cos^2 27^\circ = I_B \cos^2 (90^\circ - 27^\circ) = I_B \sin^2 27^\circ$$

$$I_A / I_B = \tan^2 27^\circ = 0.26$$

⊙

[This principle is used in intensity comparison in many instruments.]

16·2. Polarization by Reflection : Brewster's Law. Arrange a lamp, a clean glass plate (not mirror), and a receiver (eye or a photocell) for the reflected light. Let the angle of incidence be *i*. Now hold a polaroid (or tourmaline plate) in the path of the reflected light and rotate it about the beam axis. You will notice that the transmitted light changes in intensity, *i.e.*, the reflected light is not symmetrical about the ray direction. In each complete rotation of the polaroid we get two positions of maximum intensity and two of minimum intensity. The minimum intensity is, in general, not zero. The beam thus becomes *partially polarized* on simple reflection.

---

* A direction *in* a plane is not uniquely defined ; direction *normal* to a plane is unique.

If we now change the angle of incidence i and repeat the above experiment, we find that for a particular angle i_p the minima of intensity become zero. This means that the reflected beam is now a *plane polarized* beam. This particular angle of incidence i_p is known as the polarizing angle or *Brewster's angle. It is that particular angle of incidence for which the reflected beam is plane polarized.* Fig. 16·4 shows schematically the vibrations in the incident and reflected beam at the polarizing angle.

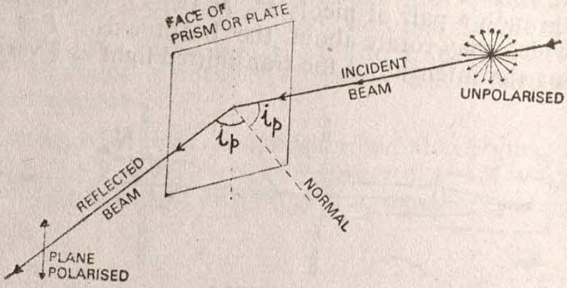


Fig. 16·4.

Polarization by reflection

Brewster determined this angle for different reflecting surfaces and established the relation

$$\tan i_p = \mu \qquad \ldots(16{\cdot}2)$$

i.e., the tangent of polarizing angle for a given material is equal to the refractive index of that material (for the light used). This is known as *Brewster's Law.*

It is interesting to note that at the polarizing angle the paths of the reflected and refracted beams are mutually at right angles. We leave the proof of this (from Eq. 16·2) as an exercise for the reader.

Note that the above statements and laws apply for reflection at the simple surface of any transparent medium, but not for reflection at silvered surfaces like mirrors.

Example 16·2. The refractive indices of glass and water are 1·54 and 1·33 respectively. Calculate the polarizing angle for a beam incident from (i) water to glass, (ii) glass to water, (iii) glass to air, (iv) air to glass.

Solution. For case (*i*), $\mu = 1{\cdot}54/1{\cdot}33 = 1{\cdot}155$

$$i_p = \tan^{-1}(1{\cdot}155) = 49°8'$$

For case (*ii*), $\mu = 1{\cdot}33/1{\cdot}54 = 0{\cdot}866$

$$i_p = \tan^{-1}({\cdot}866) = 40°52'$$

Similarly, for case (*iii*), $i_p = 33°$, case (*iv*) $i_p = 57°$ ⊙

Fresnel gave a detailed theory of reflection and refraction at the boundary of two transparent media. This theory included an explanation of polarization by reflection and also for Brewster's law.

It is interesting that Fresnel's theory was on the basis of 'elastic' waves in 'ehter'. However, the same results are now obtained on the electromagnetic theory of light ; the equations retain the same character, only the symbols change in their meanings.

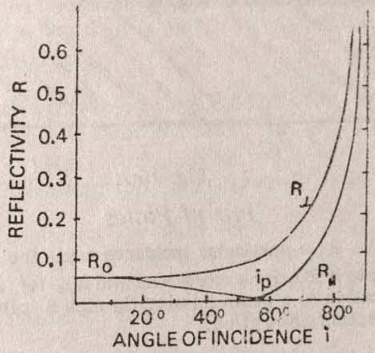


Fig. 16·5.

All the vibrations in the incident beam may be resolved into vibrations *in* and *perpendicular* to the plane of incidence. For brevity we shall call them $\parallel$ and $\perp^r$ components. Fresnel deduced the coefficient of reflection (R) for the intensity of $\parallel$ and $\perp$ vibrations separately for different angles of incidence, and got the following results, shown schematically in Fig. 16·5 :

(*i*) R for the $\perp^r$ component starts with a small value R_0 for $i=0$ and increases continuously as i increases.

(*ii*) R for the $\parallel$ component starts with the *same* value R_0, first falls and reaches zero value at a particular incidence angle $i=i_p$, and then rises, as i increases :

(*iii*) The angle i_p satisfies Brewster's Law, Eq. (16·2).

(*iv*) R_0, the reflectivity at normal in incidence, is given by

$$R_0=\left(\frac{\mu-1}{\mu+1}\right)^2 \qquad \ldots(16{\cdot}3)$$

In diagrams it is coustomary to show $\perp^r$ vibrations by dots (.) and the $\parallel$ vibrations by lines as shown in fig. 16·6. This figure shows that dots are reflected more favourably, so that the transmitted beam is left with more of the $\parallel$ vibrations.

16·3. Polarization by Refraction through "Pile of Plates".

The 'pile of plates' consists of several clean plates of glass separated by air gaps and arranged as shown in fig. 16·7, so that the beam is incident at Brewster's angle. At each plane, the refracted beam carries *all* the $\parallel$ vibrations ; only some $\perp^r$ vibrations are reflected. By suffering a large number of reflections, the transmitted beam becomes successively poorer and poorer in $\perp^r$ vibrations and yet transmitting all the $\parallel$

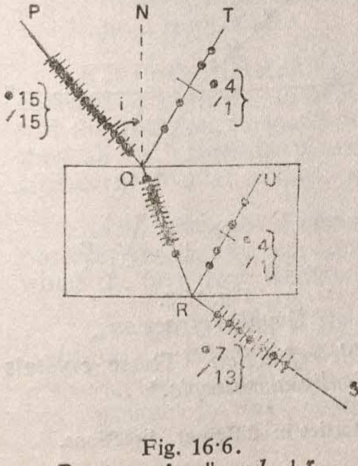


Fig. 16·6.
Representing $\parallel$ and $\perp^r$ components in a wave

vibrations. The transmitted beam is thus plane polarized with vibrations in the plane of incidence.

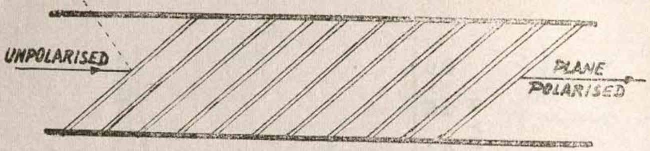


Fig. 16·7.

Pile of Plates

Example 16·3. At a particular incidence angle the cofficient of amplitude reflection of glass plate is 2% for ‖ vibrations and 8% for ⊥ vibrations. Calculate the degree of polarization of a beam passing through a pile of 10 plates at this incidence angle, neglecting absorption.

Solution. 10 plates have 20 surfaces. At each surface 98% of one and 92% of the other vibration is transmitted (in amplitude). If a is the initial amplitude in each plane (‖ and ⊥ to plane of incidence), the emergent beams have amplitudes

$$a_1=a\times(\cdot98)^{20}=\cdot667a, \qquad a_2=a\times(\cdot92)^{20}=\cdot189a$$

The corresponding intensities are (on squaring)

$$I_1=I_0\times\cdot445 \qquad I_2=I_0\times\cdot036$$

$$\text{Degree of Polarization}=\frac{I_{max}-I_{min}}{I_{max}+I_{min}}=\frac{\cdot445-\cdot036}{\cdot445+\cdot036}=0\cdot85$$ ⊙

16·4. Double Refraction. There are several crystals in which the simple laws of refraction are not followed. We find that in these crystals

(*i*) there are, in general, *two* refracted rays for each incident ray.

(*ii*) the two refracted rays may *not*, in general, lie in the plane of incidence.

(*iii*) the angles of refraction r_1, r_2 *vary* with i in such a way that, in general, neither $\sin i/\sin r_1$, nor $\sin i/\sin r_2$ remains constant, and

(*iv*) each of the two refracted rays is *plane-polarized*, the polarization in one being perpendicular to that in the other.

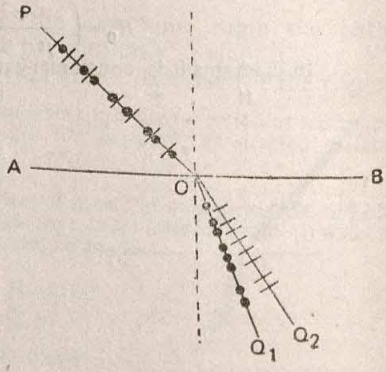


Fig. 16·8.
Double refraction

This phenomenon is called *double-refraction*. Those crystals in which this occurs are said to be *optically-anisotropic*.*

---

* Anisotropic=not having identical properties in different directions.

In some anisotropic crystalline media *one* of the two refracted rays follow the laws of refraction, only the *other* ray does not follow these laws. These crystals are called *uniaxial crystals*. The ray which follows the laws of refraction is called *ordinary*, that which does not follow the laws of refraction is called *extraordinary*. But both of them are plane polarized.

When both the refracted rays disobey the laws of refraction, the medium is called a *biaxial crystal*. But we shall limit our discussion to uniaxial crystals only. The most important example of uniaxial crystals is *Iceland spar*, also known as *calcite*, $CaCO_3$.

Fig. 16·9 (*a*) shows the view when an object is seen through a plate of calcite crystal. If the calcite plate is rotated, one image does not alter in position, but the other moves round the first one. This other image is due to the extraordinary ray.

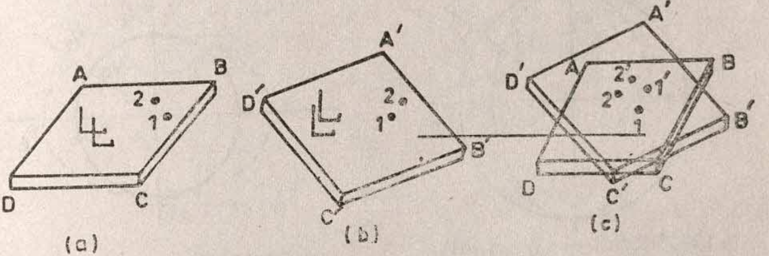


Fig. 16·9

Viewing through calcite crystals

In fig. 16·9 (*b*), let **2** be that image which rotates (round 1) if the crystal is rotated. Now if a second calcite crystal is placed over the first, four images are seen (fig. 16·9 *c*). For 1 of fig. (*a*) there are two images 1, 1′, and similarly, for 2 of fig. (*a*) there are two images 2, 2′. Further, if the second crystal is rotated, 1′ and 2′ rotate around 1 and 2 respectively.

It is further interesting to look at these four images through a polaroid or a Nicol. In one position, images 1′ and 2′ disappear with 90° rotation, images 1 and 2 disappear. Thus, the waves which show all the images are plane polarized, and the polarization for 1′, 2′ is perpendicular to that for 1, 2.

16·5. Huygens' Explanation of Double Refraction.* The above observations may be explained, using the following postulates, of which the first three were made by Huygens :

(*a*) Every point of the crystalline medium, when disturbed, sends out not one but two wavelets.

---

* This is for uniaxial crystals only. Fresnel gave a mathematical theory applicable for biaxial crystals also. It is interesting that Huygens considered longitudinal waves ; the postulate (*d*) was added by Fresnel.

(*b*) One of these wavelets expands with time as a *spherical* (*O*) wavefront and the other as a *spheroidal* (*E*) wavefront.

(*c*) The *O* (ordinary) and *E* (extraordinary) wavefronts touch along a line, called the *Optic Axis*, whose direction is fixed relative to the crystallographic axes. In some crystals the *O* wave-surface is outside ; these are called *positive* crystals ($V_0 > V_E$). In others the *O* wave-surface is inside ; these are called *negative* crystals ($V_0 < V_E$).

(*d*) Any given part of a wave-surface can carry vibrations in one direction only (plane polarized), and for a given wave-normal, the direction of vibration in one wave is perpendicular to that in the other wave.*

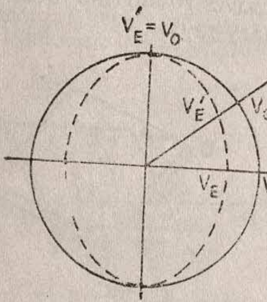


Fig. 16·10 (*a*).

Positive crystal, $V_E < V_0$
Examples : *Quartz, ZnO*

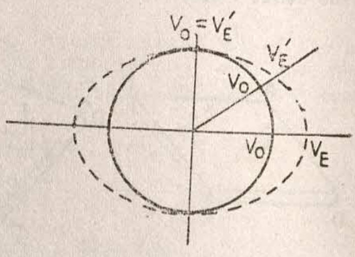


Fig. 16·10 (*b*).

Negative crystal, $V_E > V_0$
Example : *Calcite*

Fig. 16·10 shows the wave-surface at unit time. For any direction the wave-velccities are V_0 and V_E'. V_E' equals V_0 along the optic axis ; its other extreme value V_E occurs in a plane perpendicular to the optic axis and is called the extraordinary wave velocity.

We now define the 'ordinary' and 'extraordinary' refractive indices by the relations†

$$\mu_0 = c/V_0 \quad ; \quad \mu_E = c/V_E \qquad \ldots(16{\cdot}4)$$

where *c* is the speed of light in vacuum.

In § 10·3 we had used Huygens' principle to deduce the refracted ray corresponding to an incident ray for ordinary media. Huygens used the same procedure for uniaxial crystals, in which *two secondary wavelets* arise from each disturbed point on the surface of the crystal. The growth of the extraordinary wavelet depends upon how the surface of the crystal is cut relative to the optic axis. We shall consider just two cases.

---

* It may be noticed that all the four statements are direct assumptions. A theory based on ether was given by Fresnel, and one based on the *EM* nature of light and crystal structure is now available. (See § 16·6).

† For a given medium, μ_0 and μ_E are constants, but c/V_E' is not constant. It varies in the range μ_0 to μ_E.

Fig. 16·11 (*a*) shows the case of a positive crystal with optic axis normal to its face *SS*. For oblique incidence *PA* we get two refracted wave-fronts BW_0 and BW_E, travelling in different directions AQ_0 and AQ_E, respectively, with different speeds. In fact the extraordinray AQ_E does not coincide with the wave normal. For this reason the *ray velocity* is different from the *wave velocity* for the *E* wave-front.* (In the expression $\mu_E = c/V'_E$ the wave velocity V'_E is involved.) For normal incidence the *O* and *E* wavefronts would coincide and also travel with equal speeds

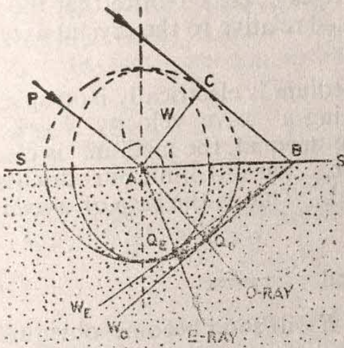
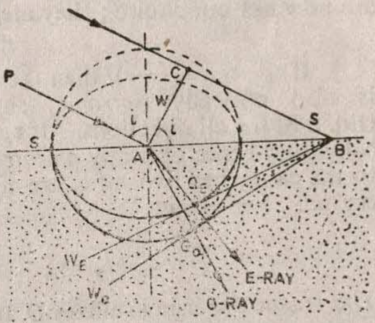


Fig. 16·11 (*a*)

Optic axis normal to surface of separation



Fig· 16·11 (*b*)

Optic axis in the surfaee of separation and in plane of incidence

Fig. 16·11 (*b*) shows the case when the optic axis is in the surface of the crystal and in the plane of incidence. For oblique incidence we again have two refracted wavefronts travelling in different directions. For normal incidence the two wavefronts W_0, W_E will now be parallel, but travel with different speeds V_0 and V_E.

16·6. Electromagnetic Theory of Double Refraction. The most general relations between the electric field **E** and dielectric displacement **D** in a crystalline medium are

$$D_x = \varepsilon_{xx} E_x + \varepsilon_{xy} E_y + \varepsilon_{xz} E_z,$$

and two similar ones for D_y and D_z.

---

* Let a straight file of soldiers march at speed V at angle θ from the line perpendicular to the file. Then the file advances along the normal at speed $V' = V \cos \theta$. The situation is similar here.

But energy conservation requires that in every case you have three mutually perpendicular axes, *fixed relative to the crystal axes*, such that the coefficients ε_{ij} are zero for $i \neq j$. Using these axes we can write

$$D_x = \varepsilon_x E_x \ ; \ D_y = \varepsilon_y E_y \ ; \ D_z = \varepsilon_z E_z$$

where single suffixes are now used ($\varepsilon_x \equiv \varepsilon_{xx}$ etc.) Notice that the axes are now not our choice ; they are fixed relative to the crystal axes.

If $\varepsilon_x = \varepsilon_y = \varepsilon_z$ $(= \varepsilon)$, then the medium is electrically isotropic. It is also optically *isotropic*, showing a wave velocity $V = 1/\sqrt{\mu\varepsilon}$ (Eq. 8·8) for all directions. If ε_i are unequal, the medium is electrically and hence optically *anisotropic*. We will limit our discussion to the case when two of ε_i are equal. That is the case of a *uniaxial crystal*. Let us put

$$\varepsilon_x = \varepsilon_y = \varepsilon_0, \qquad \text{and} \qquad \varepsilon_z = \varepsilon_E \qquad \ldots(16{\cdot}5)$$

The reason for using suffixes O and E will soon become obvious.

Consider a plane wave with wave-normal in the x direction. The E vector will be in the yz plane and let us call it E_{yz}. Then we have

$$D = \varepsilon_0 E_y + \varepsilon_E E_z$$

The D lies in the yz plane, but its direction does *not* agree, in general, with E_{yz}. So the medium would not transmit a wave with the general electric vector E_{yz}. Its component $E_1 = E_y$ will travel with speed $1/\sqrt{\mu\varepsilon_0} = V_0$, and component $E_2 = E_z$ with speed $1/\sqrt{\mu\varepsilon_E} = V_E$. Thus there are two waves, each being plane-polarised. In the same way, a wave with wave-normal along y direction will split into polarized waves with $E_1 = E_x$ and $E_2 = E_z$, transmitted with velocities V_0 and V_E respectively. The wave with wave-normal along z direction has a special situation. E has components E_x and E_y, so that $D = \varepsilon_0 E_x + \varepsilon_0 E_y = \varepsilon_0 E$. In this case directions of D and E agree for all E_x, E_y. So the wave does not split into polarized components and travels on with velocity V_o. This z direction of wave-normal represents the *optic axis*.

Let us also consider a general direction of the wave normal, say, in the xz plane. Fig. 16·12 shows the situation. The E vector (not shown) may be resolved into a component E_y perpendicular to xz plane and one in the xz plane (E_{xz}). The first gives a polarized wave travelling with velocity V_0. For the second, we note that $\varepsilon_x \neq \varepsilon_z$, and hence D_{xz} will not coincide with E_{xz} in direction. Component of D_{xz} along the wave-normal (*i e.*, the longitudinal one) will not contribute to wave-propagation, and the component transverse to the wave-normal (*i.e.*, along E_{xz}) alone will contribute. Calling this component D'_{xz}, we now see that the magnitude ratio D'_{xz}/E_{xz} is neither ε_0 nor ε_E, but is in between these extremes. Thus as angle θ (fig. 16·12) is varied from 0 to $\pi/2$ the velocity V of wave with E_{xz} polarization varies from $c/\sqrt{\mu\varepsilon_E}$ to $c/\sqrt{\mu\varepsilon_o}$. In fact one can establish that

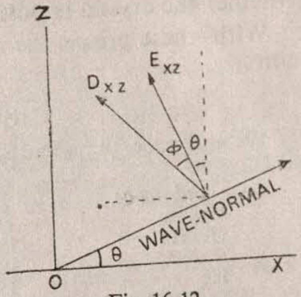


Fig. 16.12.

To deduce the extraordinary wave-velocity in a uniaxial crystal

$$\frac{D'_{xz}}{E_{xz}}=\varepsilon_E \cos^2\theta+\varepsilon_0 \sin^2\theta \qquad \text{...(16·6 } a)$$

The left side is ε', and taking $V=1/\sqrt{\mu\varepsilon'}=c/\sqrt{\varepsilon'}$ leads o

$$\frac{V^2 \cos^2\theta}{V_E^2}+\frac{V \sin^2\theta}{V_0^2}=1 \qquad \text{...(16·6 } b)$$

This is the ellipse, as proposed by Huygens (Fig. 16·10).

Thus the electromagnetic theory gives a complete and quantitative explanation of double refraction, and the plane-polarized character of the components. It is notable that the quantities ε_0 and ε_E can be determined by electrical experiments. Only the high-frequency values are applicable to optics.

16.7. Experimental Determination of μ_0 and μ_E. In terms of the Snell's equation $\sin i/\sin r=\mu$ we can determine μ_0 alright by the prismatic deviation method, but for the extraordinary ray this ratio has, in general, no meaning since it varies widely with i. If, however, it is so arranged that the refracted ray travels in the crystalline medium in a direction perpendicular to the optic axis, then $\sin i/\sin r_E$ gives us μ_E.

Hence, for determination of μ_0 and μ_E by a prism experiment we choose a prism which is so cut from the crystal that the optic axis is either (*i*) normal to the base of the prism, or (*ii*) parallel to the refracting edge of the prism as shown schematically at (*a*) and (*b*) respectively in Fig. 16·13. In each case the rays *inside* the prism travel perpendicular to the optic axis. If the vibrations are in the plane of paper in the upper ray and perpendicular to it in the lower

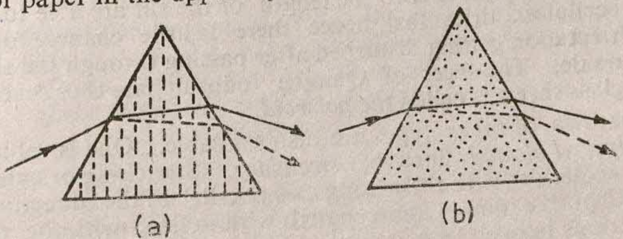

(a) (b)

Fig. 16·13.

Prisms for determining μ_0 and μ_E. Dotted lines or dots indicate the optic axis ray in fig. (*a*), it is other way in fig. (*b*). The actual thing depends on whether the crystal is positive ($V_0 > V_E$) or negative ($V_0 < V_E$).

With these prisms the refractive indices are given by the usual equation

$$\mu = \frac{\sin \frac{1}{2}(A+\delta)}{\sin \frac{1}{2} A}$$

We give below the μ_0 and μ_E values (for Sodium *D* light) for some typical uniaxial crystals

Substance	μ_0	μ_E	$\mu_E - \mu_0$
Calcite	1·658	1·486	—0·172
Quartz	1·5442	1·5534	+0·0092
NaNO$_3$	1·5874	1·3361	—0·2513
Ice	1·3091	1·3104	+0·0013
Tourmaline	1·659	1·638	—0·021

Note that the difference of μ_0 and μ_E is very large for calcite and sodium nitrate crystals, *i.e.*, these crystals are strongly anisotropic.

Example 16·4. Calculate the angular deviations of *E* and *O* rays passing through a 60° prism of NaNO$_3$ in minimum deviation condition, optic axis being perpendicular to the base.

Solution. Using the equation above with $A = 60°$,

$$\sin \tfrac{1}{2}(60+\delta_0) = 1{\cdot}587 \times 0{\cdot}5 = 0{\cdot}794 \Rightarrow \delta_0 = 45° \; 12'$$

$$\sin \tfrac{1}{2}(60+\delta_E) = 1{\cdot}336 \times 0{\cdot}5 = 0{\cdot}668 \Rightarrow \delta_E = 23° 48'$$

[Note that the difference between μ_0 and μ_E is much larger than the variations of μ with wavelength in most cases. For extra-dense-flint glass difference of μ from violet to red light is less than 0·03.]

16·8. The Polaroids. Certain double refracting crystals not only split an incident unpolarized beam of light into two mutually perpendicular plane-polarized beams, but also absorb one of the beams as it travels in them, while passing the other without much absorption. The behaviour is called *dichroism*, and tourmaline is one example of a dichroic crystal. A plate cut from a dichroic crystal, therefore, produces plane polarized light by transmission.

To get large polarizing plate on this principle a special process was developed. *Iodosulphate of quinine* shows dichroism but its crystals are very small. We have to arrange a large mass of these microcrystals with their optic axes all parallel. In one such process a paste of these crystals is prepared in *nitrocellulose*, it is squeezed past a fine slit and on the other side it is received between two thin glass sheets, where it dries and gets fixed. Quinine Iodosulphate has a long molecule and in squeezing past the slit only those molecules whose axes become parallel to the length of the slit are able to pass. The nitrocellulose dries fast, hence there is little chance for the regular orientation getting disturbed after passing through the slit on the other side. The sheet of Quinine Iodosulphate thus enclosed between glass sheets is called *the polaroid.*

Polaroids of other types are also available. One is obtained from *polyvinyl alcohol* films by *stretching*. Stretching orients the complex molecules with their long axis parallel to the direction of stress. Then the film is impregnated with iodine, with the result that vibrations in one plane are absorbed selectively, while those in a perpendicular plane are passed well. This is called *H polaroid*. If instead of iodine impregnation, the stretched film is heated with a dehydrating catalyst, dichroism is again developed. This is called *K polaroid.*

The absorption in the polaroids arises from the high electric conductivity provided by the iodine atoms along the length of the long-chain molecules. The arrangement is equivalent to a wire-grid polariser (§ 16·1) and that component of the wave which has electric field vibrations parallel to the long chain of themolecules is absorbed, while the other component is passed through. In an ideal case we expect 50% transmission, but a typical transmission is around 35% only. The extra losses are partly due to reflection at the two surfaces, inside the material.

Polaroid sheets may be fitted in car headlights and wind screens to avoid glare. If the headlight sheet allows vibrations at 45° clockwise from the vertical, and the windscreen sheet also allows vibrations at 45° clockwise from vertical, then a driver of vehicle opposite will not be dazzled by your headlight, nor will you be dazzled by the headlight of the vehicle opposite you.

If two polaroids are put one behind the other on a window, and one of these can be rotated, then the amount of transmitted light can be controlled easily. Use is made of this in big hotels, aeroplanes, etc.

Sunlight reflected from a horizontal surface (like ice or lake water) has more of the horizontal vibrations than the vertical ones. If your sunglasses are of polaroids so arranged as to cut horizontal vibrations, then glare will be considerably reduced.

16.9. Nicol Prism. This is made from a natural crystal of *calcite* ($CaCO_3$ hydrated), also called *iceland spar.* It is a slant prism on

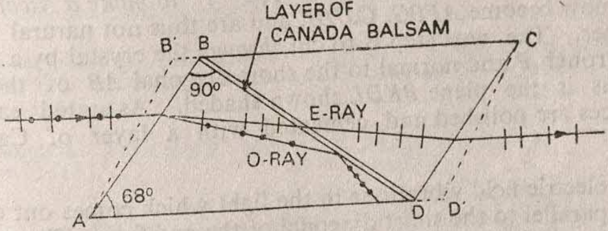


Fig. 16·14.

Principal section of a nicol prism. The natural end faces AB', CD' are cut off to AB, CD with the purpose that for a cone of rays of 14° angle, with chief ray parallel to length of the prism, the condition for total internal reflection of O-rays is satisfied.

a rhombus base (Fig. 16·14). We cut a piece with length *AD* three times the edge *AB*. The crystal is cut along *BD* in the manner described below and after polishing the cut surfaces to optical flatness they are cemented with a layer of *Canada balsam.* Calcite is a negative crystal ($\mu_0 > \mu_E$) and Canada balsam has a refractive index μ_0 lying between μ_0 and μ_E. When an incident ray parallel to the length of the prism falls on one face, it breaks into *O* and *E*-rays. When these reach the Canada balsam layer, for the *O*-ray this layer has lower refractive index than calcite and the angle of incidence is greater than $\sin^{-1}(\mu_c/\mu_0)$. Hence the *O*-ray suffers total internal reflection.* The purpose in choosing the length 3 times the side of a face is just to satisfy this condition. The *E*-ray alone passes through the prism, and thus we get plane-polarized light from an incident unpolarized beam.

The manner of cutting the calcite to get a Nicol will now be described. Every crystal has certain natural faces along which it can be easily 'cleaved', and the angles between natural edges are characteristic. In calcite the most important edges are $B'F'$, $B'G'$, $B'C$. As regards lengths, for a Nicol we choose $B'F'=B'G'$ and $B'C \approx 3B'F$.

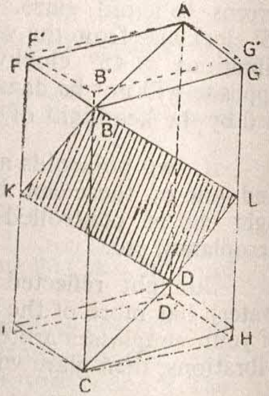


Fig. 16·15.
Cutting a calcite prism to make a Nicol

Now, in this crystal the three edges at B' (and at the diagonal corner D') meet in obtuse angles, which are all equal. B' and D' are called the *blunt corners.* The *optic axis* in calcite makes equal angles with the edges meeting at B' (or D'). AB' and CD' are *shorter diagonals* of the end faces.

In preparing a Nicol, the first step is to cut the end faces such that angles $B'AD'$ and $B'CD'$, which are 71° in the natural crystal, are reduced to 68°. The end faces now become *AFBG*, *CJDH*, and are thus not natural cleavage planes. The next step is to cut *through* the crystal by a plane passing through *B* and normal to the short diagonal *AB* of the cut face. This is the plane *BKDL* shown shaded. As stated earlier, the cut faces are polished and cemented with a layer of Canada balsam.

The electric field vibrations in the light which comes out of the Nicol are parallel to the short diagonal of the end face. The section *ABCD* (through the short diagonals) is called the *Principal Section*, and *E*-field vibrations in the emergent ray lie in the principal section.

---

* The values of μ_0, μ_E and μ_c for sodium yellow are as follows :

$$\mu_0=1{\cdot}658, \quad \mu_E=1{\cdot}486, \quad \mu_c=1{\cdot}552$$

The critical angle for total reflection of *O*-ray is, therefore,

$$\sin^{-1}\frac{\mu_c}{\mu_0}=\sin^{-1}\frac{1{\cdot}552}{1{\cdot}658}=69°.$$

16·10. Double Image Prisms. These are prisms of double refracting uniaxial crystals made with the purpose of getting both the plane polarised beams out with an angular separation. One such is shown in Fig. 16·16. It has two right angled prisms *A* and *B*, such that the incident beam *PQ*, the optic axis in A and that in B are all mutually perpendicular.

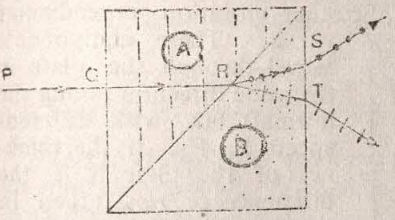


Fig. 16·16.
A double image prism

The vibrations in the unpolarized incident beam *PQ*, on entering prism *A*, break into vibrations || and ⊥ʳ to the plane of paper. Both these plane-polarized vibrations travel along the same line *QR*, although one travels with speed V_0 the other with speed V_E. On entering the second prism *B*, the *O*-vibration becomes *E* and *E*-vibration becomes *O*,* hence the speeds interchange. In other words, the refractive index of *B* with respect to *A* is V_0/V_E for one ray and V_E/V_0 for the other ray. Therefore, the two plane-polarized beams are separated, as shown by *RS*, *RT*.

A typical instrument using a double image prism is a spectrophotometer, which measures the absorption by a medium at different wavelengths. White light from a slit is passed through a double image prism ; the separated beams travel through two equal paths, one of which contains the medium under test. On emergence both the beams are focussed to form adjacent images along the length of a spectrometer slit. In the telescope of the spectrometer we get two spectra in juxtaposition. We see it through a polaroid. The polaroid adjustment needed to match intensities at a given λ gives a measure of the absorption by the test medium at that λ. [See Example 16·1.]

Example 16·5. Let a double image prism like that of Fig. 16·16 be made of a material for which $\mu_0 = 1{\cdot}662$, $\mu_E = 1{\cdot}474$. Deduce the angular separation of the emergent beams.

Solution. The angle of incidence at *R* (Fig. 16·16) is 45°. The angles of refraction are :

$$\sin^{-1}\left(\frac{1}{\sqrt{2}} \times \frac{1{\cdot}662}{1{\cdot}474}\right) = 52^\circ 50' \; ; \; \sin^{-1}\left(\frac{1}{\sqrt{2}} \times \frac{1{\cdot}474}{1{\cdot}662}\right) = 38^\circ 50'$$

These rays fall on the opposite face at angles of incidence $+7^\circ\ 50'$ and $-6^\circ\ 10'$ respectively. The angles of emergence are

$$\sin^{-1}(1{\cdot}474 \sin 7^\circ\ 50') = 11^\circ\ 34'$$

$$\sin^{-1}(1{\cdot}662 \sin -6^\circ\ 10') = -10^\circ\ 12'$$

The angle between the emergent beams is therefore 21° 46′. ⊙

16·11. Phase Retardation Plates. These are plates of uniaxial crystals cut with optic axis parallel to the faces of the plate (Fig. 16·17). When a *plane polarized* beam is incident normally on such

* This is because the optic axis of *B* is ⊥ʳ to that of *A*,

a plate, the vibrations in the beam break at the first face into one vibration along the optic axis and the other vibration perpendicular to it. These components travel through the plate in the same direction (along the normal) but with different speeds V_0, V_E. If the thickness of the plate is d, the difference of *time* taken in traversing the plate by the two beams is

$$|\,\Delta t\,| = d.\left(\frac{1}{V_o} - \frac{1}{V_e}\right)$$

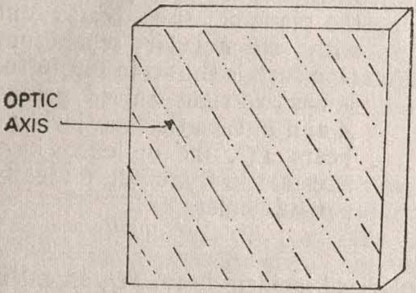


Fig. 16·17.
A phase retardation plate

In terms of equivalent length in vacuum, the relative retardation is

$$\Delta x = c\Delta t = d\,|\,\mu_o - \mu_e\,|$$

Thus a relative phase retardation $(2\pi/\lambda)\Delta x$ is created between the two mutually perpendicular and coherent* vibrations. For this reason the plates are called *phase retardation plates*. We may choose the thickness to make the retardation in path equal to $\lambda/4$ or $\lambda/2$; the plate then becomes *quarter-wave plate* or *half-wave plate* respectively. If the path retardation is $(n \pm \frac{1}{4})\lambda$, with n as any integer, then also the effect is the same as due to $\frac{1}{4}\lambda$ retardation.† Thus, for a quarterwave plate we require $\Delta x = (n \pm \frac{1}{4})\lambda$, where λ is the wavelength in vacuum. Hence

$$(n \pm \tfrac{1}{4})\lambda = d\,|\,\mu_o - \mu_E\,| \qquad \text{for QW plate}\ldots(16{\cdot}7a)$$

For a half-wave plate, similarly

$$(n \pm \tfrac{1}{2})\lambda = d\,|\,\mu_o - \mu_E\,| \qquad \text{for HW plate}\ldots(16{\cdot}7b)$$

In practice the QW or HW plate has to be supported between glass plate (see Ex. 16·6).

Example 16·6. For calcite, $\mu_0 = 1{\cdot}658$, $\mu_E = 1{\cdot}486$ for sodium yellow. Calculate the thickness of the thinnest *QW* plate of calcite for sodium yellow.

Solution. $d = \dfrac{5893 \times 10^{-8}\text{cm}}{4(1{\cdot}658 - 1{\cdot}486)} = \dfrac{5893}{4 \times 1{\cdot}72} \times 10^{-8} = {\cdot}86 \times 10^{-4}\text{cm}.$

[Obviously such a thin plate will have to be supported between glass plates].

⊙

Example 16·7. A quarter-wave plate is meant for $\lambda_0 = 5{\cdot}893 \times 10^{-5}$ cm. What phase retardation ϕ will it show for $\lambda = 4{\cdot}358 \times 10^{-5}$cm ? (Neglect changes of μ_o and μ_E with λ).

---

* They are coherent because both of them are components of the same plane polarized beam.

† If retardation $(n + \frac{1}{2})\lambda$ causes a phase *lag* $\pi/2$, retardation $(n - \frac{1}{2})\lambda$ causes phase *advance* $\pi/2$ in one vibration relative to the other perpendicular coherent vibration.

Solution. Assuming that *QW* plate is of the smallest thickness.

$$|\mu_0-\mu_E|\,d=\tfrac{1}{4}\times 5{\cdot}893\times 10^{-5}\text{ cm}$$

Phase retardation for $\lambda=4{\cdot}358\times 10^{-5}$ cm

$$\phi=\frac{2\pi}{4{\cdot}358\times 10^{-5}}\times\frac{5{\cdot}829\times 10^{-5}}{4}=0{\cdot}67\pi\ \textbf{radian}.$$

[Thus a *QW* plate for yellow is *not* a QW for other colours]. ⊙

16·12. Elliptically and Circularly Polarized Light.

Consider two *coherent* waves travelling in z-direction, and polarized in mutually perpendicular directions :

$$E_x=a\cos(\omega t-kz)\;;\;E_y=b\cos(\omega t-kz+\phi) \qquad \ldots(16{\cdot}8)$$

These are coherent if phase difference ϕ is independent of time.

The superposition of two such waves makes the E vector rotate in the x-y plane, with its tip following an elliptical path. We may see this for $\phi=\pm\pi/2$ by writing the coordinates of the tip of E vector at $z=0$:

$$x=a\cos\omega t\;;\quad y=b\cos(\omega t\pm\pi/2) \qquad \ldots(16{\cdot}9)$$

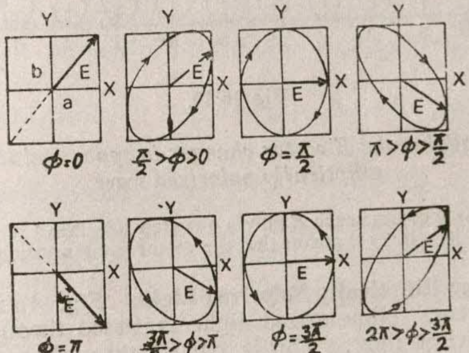


Fig. 16·18.

Oscillations $E_x=a\cos wt$ and $E_y=b\cos(\omega t+\phi)$ producing an E-vector, whose time variations depends on ϕ. Thick arrow shows E at $t=0$. E rotates clockwise for $\pi>\phi>0$, and antickwise for $2\pi>\phi>\pi$.

Elimination of t between them gives

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \qquad \ldots(16{\cdot}10)$$

which is an ellipse.

The direction of rotation of E is settled by considering the coordinates at $t=0$ and a little later when ωt is very small. Eqs. (16·9) gives the coordinates as $(a, 0)$ and $(\sim a, \pm\omega t)$. Thus, for $\phi=+\pi/2$ in Eq. (16·9) the vector makes clockwise rotation as seen by the observer. For $\phi=\pi/2$ the rotation is anticlockwise.

Fig. 16·18 shows more generally how the E vector changes with time for various ϕ values. With $\phi=0$ and $\phi=\pi$ the resultant wave remains plane polarized, but with the plane of polorization making angles $\tan^{-1}(b/a)$ and $\tan^{-1}(-b/a)$ respectively with the E vibration.

If the E_x and E_y waves have equal amplitudes $(a=b)$, then for $\phi=\pm\pi/2$ the wave becomes circularly polarized.

The changes in E vector in an elliptically or circularly polarized wave may be appreciated from Fig. 16·19. Ellipses are shown at a number of equidistant points along the line of sight. Each ellipse is divided into eight equal sectors. The thick line represents the E vector at a given instant of time. The wavelength λ corresponds to the smallest separation for which E vector repeats in value. Notice that value here means vector value ; in fact in circularly polarized magnitude of E does not change, only the direction changes.

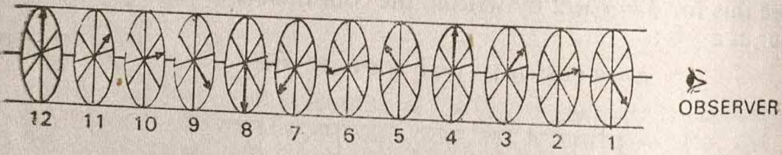


Fig 16·19.

For visualizing the E vector changes in space and time in an elliptically polarized wave

With passage of time each E moves one step $\lambda/8$ ahead in time $T/8$. For the observer in Fig. 16·19 it means that the vector rotates anticlockwise.

Producing Elliptically Polorized Light. Fig. 16·20 (*i*) shows the arrangement. Unpolarized light is passed through a Nicol to get plane polarized light. This is passed through a quarter wave plate*, which does two things : (i) breaking the polarized wave of amplitude a_0 into two mutually perpendicular components of amplitudes $a=a_0\cos\theta$ and $b=a_0\sin\theta$ as Fig. 16·20 shows and (ii) producing a phase difference $\pi/2$ between the two components as they travel through the plate. Thus we have two emergent waves

$$E_x=a\cos(\omega t-kz) \qquad E_y=b\cos(\omega t-kz\pm\pi/2)$$

The result is an elliptically polarized wave, as discussed. The angle θ between the short diagonal of the Nicol and optic axis of the QW plate determines the relative magnitudes of the major and minor axes of the ellipse, *i.e.*, $|E|_{max}$ and $|E|_{min}$.

---

* Any phase retardation plate would do, except one with ϕ near $n\pi$. QW plate is convenient, because major and minor axes of the ellipse are then known.

† In the case of light, *coherence* can be obtained if both the perpendicularly polarized waves are obtained from a single plane polarized wave.

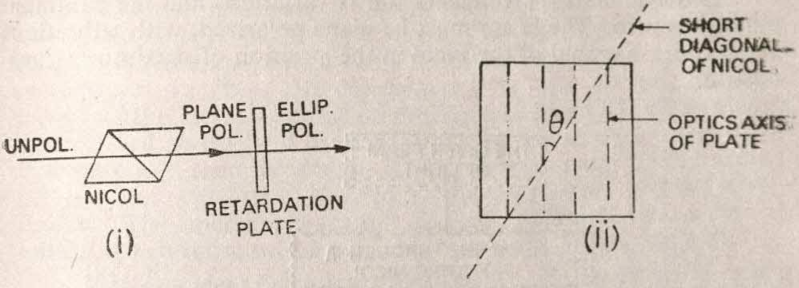


Fig 16·20.

(*i*) *Setup for producing elliptically or circularly polarized light,*
(*ii*) *Angular relation between axes of the nicol and the retardation plate*

Producing Circularly Polarized Light. The same arrangement of Fig. 16·20 with $\theta = 45°$ or 135° produces circularly polarized. The direction of rotation will be opposite in the two cases.

16·13. Analysis of Polarization in a Given Beam of Light. From the point of view of polarization a given beam of light may be :

(*a*) Unpolarized,
(*b*) Partially plane-polarized,
(*c*) Plane-polarized,
(*d*) Circularly polarized, or
(*e*) Elliptically polarized.

where we do not consider mixtures. Before we proceed to describe the experimental procedure for anlysis, let us note that

(*i*) There are two cases of complete symmetry, *viz.*, unpolarized light and partially polarized light.

(*ii*) There are two cases of partial asymmetry, *viz.*, elliptically polarized light and partially polarized light.

(*iii*) There is one case of complete asymmetry : plane polarized light.

Let us now apply the tests and analyze the observations :

Test A : *Pass the beam through a nicol and note the changes in intensity as the nicol is rotated.*

Observation (*i*) : No change of intensity whatsoever. The beam must be either unpolarized or circularly polarized.

Observation (*ii*) : Intensity shows variations—two maxima and two minima in one rotation—but minimum intensity is not zero. The beam must be partially plane polarized or elliptically polarized. The short diagonal of nicol in the position of maximum intensity gives us the major axis of the elliptical vibration or the direction of maximum vibrations in the partially polarized beam.

Observation (iii) : Intensity shows variations and the minimum intensty is zero. The beam must be plane polarized, with vibrations along short diagonal of the nicol in the position of maximum transmisson.

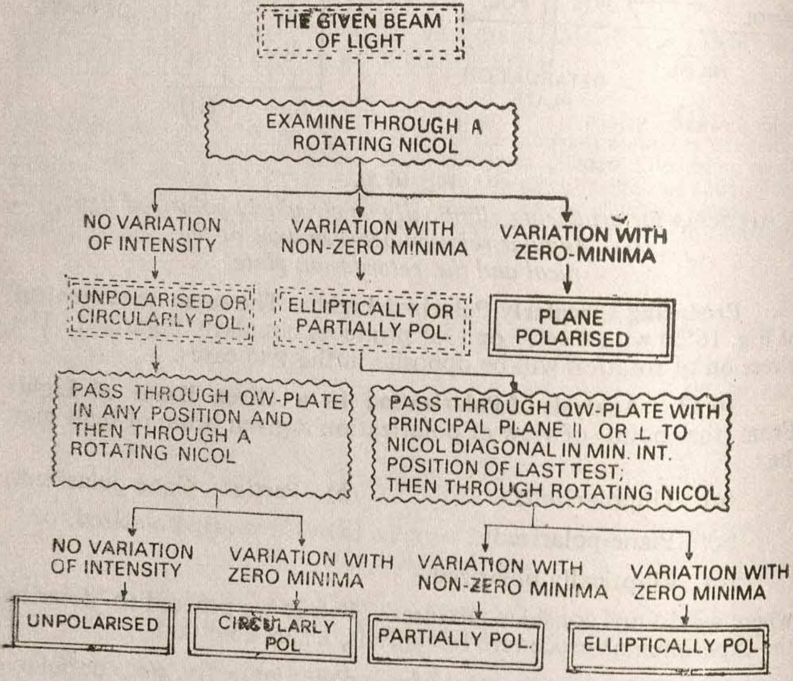

THE GIVEN BEAM OF LIGHT
EXAMINE THROUGH A ROTATING NICOL
NO VARIATION OF INTENSITY
VARIATION WITH NON-ZERO MINIMA
VARIATION WITH ZERO-MINIMA
UNPOLARISED OR CIRCULARLY POL.
ELLIPTICALLY OR PARTIALLY POL.
PLANE POLARISED
PASS THROUGH QW-PLATE IN ANY POSITION AND THEN THROUGH A ROTATING NICOL
PASS THROUGH QW-PLATE WITH PRINCIPAL PLANE ∥ OR ⊥ TO NICOL DIAGONAL IN MIN. INT. POSITION OF LAST TEST; THEN THROUGH ROTATING NICOL
NO VARIATION OF INTENSITY
VARIATION WITH ZERO MINIMA
VARIATION WITH NON-ZERO MINIMA
VARIATION WITH ZERO MINIMA
UNPOLARISED
CIRCULARLY POL
PARTIALLY POL.
ELLIPTICALLY POL

Fig. 16.21.

Analysis of polarized light

Test B : To settle between circularly polarized and unpolarized light, *pass the beam first through a quarter-wave plate and then examine it through a nicol.* If the beam is circularly polarised, the *QW* plate will break it into two $\perp^r$ vibrations of phase difference $\pi/2$, them in passing through the plate a further phase difference $\pi/2$ or $-\pi/2$ will be created between them, so that the two emergent $\perp^r$ vibrations will have phase difference 0 or π. In either case, their resultant will be a plane-polarized vibration. So, on rotating the analyzing nicol, there will be two positions of zero intensity. This will show that the incident beam is circularly polarized.

If, however, the original beam is unpolarized, them the infinite number of plane vibrations of all orientations will be converted by the *QW* plate into an infinite set of elliptical vibrations of all orientations, and an examination through a rotating nicol will show no change of intensity.

Test C : To settle between elliptically polarized and partially plane polarized light, *pass the beam first through a QW plate, so placed that its optic axis lies parallel or perpendicular to that position of short diagonal of nicol where it showed minimum intensity in the previous test A. Then examine the emergent light through a rotating nicol.* If the incident vibration is elliptical, then *QW* plate (placed as described above) will break it into two $\perp^r$ vibrations of phase difference $\pi/2$, then in passing through the *QW* plate a phase difference $+\pi/2$ or $-\pi/2$ will be added, so the net phase difference will be 0 of π. The emergent beam will be plane polarized ; examined by a nicol it will show two zero-intensity minima in each rotation.

If the incident beam is partially polarized, each of its vibrations will be converted by the *QW* plate into an elliptical vibration and examination through a nicol will show vibration of intensity with a non-zero minimum.

Summary of the analysis is given in Fig. 16·21.

PROBLEMS

16·1. Show that for incidence at Brewster's angle, the angle between the incident and refracted rays is 90°.

16·2. Sunlight reflected from still water in a lake is observed through a polaroid. When the polaroid is rotated in its own plane the intensity changes. Explain.

16·3. Calculate the polarizing angle for an air-diamond interface when light is incident (*i*) from air, (*ii*) from diamond. (Given n of diamond$=1{\cdot}80$).

16·4. Deduce Eqs. (16·6a) and (16·6b) for the e.m. theory of double refraction in a biaxial crystal.

16·5. A biaxial crystal has wave-velocities V_o and V_E. The optic axis is designated as x-axis. For a wave with wave-normal along the z-axis, deduce the velocities and directions of E vectors for the two polarized waves possible.

16·6. Repeat Problem 16·5 with wave-normal in the xy plane, making angle θ with the y-axis.

16·7. A 'polarizer' is that which would convert ordinary unpolarized light to plane polarized light. The ideal polarizer is that which would give 50% of the original light as plane polarized one. Discuss this 50% business.

16·8. A beam of light is passed through a polaroid. On rotating the polaroid the intensity varies, the maxima occurring at $\theta=12°$ and 192°. The minima are of zero intensity. Deduce (*i*) the angular positions of the minima, and (*ii*) I/I_{max} at positions $\theta=42°$ and $\theta=72°$.

16·9. Define 'ordinary' and 'extraordinary' refractive indices for a uniaxial anisotropic material. For the extraordinary beam the ray and wave normal generally do not coincide. Explain.

16·10. In some photometers two side-by-side beam *A* and *B* are obtained, plane polarized in mutually perpendicular directions. In a particular case these were observed through a polaroid. From a setting where *A* showed zero intensity, the polaroid had to be turned by 18° to get *A* and *B* equalized in intensity. Deduce the intensity ratio I_A/I_B. Also deduce the ratio of *observed* intensities for 2° error in the setting.

16·11. Explain the action of a tourmaline plane in producing plane polarized light. What will be the polarization of the transmitted light if the incident light has (*i*) circular polarization, (*ii*) elliptical polarization ?

16·12. Explain the possible polarizations you can produce from an unpolarised beam using combinations of (*i*) a nicol and a *QW* plate, (*ii*) a nicol and a *HW* plate.

16·13. A transparent plate is given. Using two nicols (or polaroids) how would you find whether the plate is a *QW* plate, a *HW* plate or a simple glass plate.

16·14. In making a nicol prism what are the reasons for (*i*) having the length$\sim$3 times the edge width, (*ii*) cutting off the end faces to reduce the angle from 71° to 68°, (*iii*) using Canada balsam layer, and (*iv*) polishing the cut faces before applying Canada balsam ?

16·15. For calcite $\mu_o=1{\cdot}659$, $\mu_e=1{\cdot}488$ for $\lambda=5461$ *A*. Calculate the possible thickness of a plate which would act as (*i*) a half-wave plate, (*ii*) quarter wave plate for this light.

16·16. A sheet of cellophane is so made that it is a half wave plate for light of $\lambda=4200$ *A*. How would it act for light of $\lambda=8400$ *A*, neglecting change of $\mu_o-\mu_E$ with λ. Would this result always hold ?

16·17. A half-wave plate is introduced in the path of a plane-polarized beam. Discuss the emergent beam for various orientations of the half-wave plate.

16·18. A phase retardation plate of quartz has thickness 0·1436 mm. For what wavelengths in the visible region ($\lambda=4500$ to 8000 *A*) will it act as (*i*) a *QW* plate, (*ii*) a *HW* plate ? Given : $\mu_o=1{\cdot}5443$ and $\mu_e=1{\cdot}5533$, difference only slightly depending on λ.

16·19. From an unpolarized beam two plane-polarized beams are obtained with a double-image prism. These are later made to travel along the same direction, with path retardation $\lambda/2$ in one path. Experiment shows that they do *not* give us an elliptically polarized beam. Explain.

16·20. Consider the following waves :

(*a*) $E_x = E_2 \cos(\omega t - kz)$, $E_y = O$

(*b*) $E_x = E_2 \cos(\omega t - kz)$, $E_y = E_2 \cos\left(\omega t - kz - \frac{\pi}{2}\right)$

(*c*) $E_x = E_3 \cos(\omega t + kz)$, $E_y = E_3 \cos\left(\omega t + kz + \frac{\pi}{2}\right)$

(*d*) $E_x = E_4 \cos(\omega t - kz)$, $E_y = E_5 \cos(\omega t - kz)$

Now answer the following—

(*i*) Which of these is plane-polarized, circularly polarized, or elliptically polarized ?

(*ii*) Which of these has right-handed rotation, and which a left handed one ?

(*iii*) What will be the result of superposition of (*a*)+(*b*) ; of (*a*)+(*d*) ?

17

Rotation of Plane of Polarization

17·1. Optical Rotation. In Fig. 17·1, let P represent a polarizer and Q an analyzer (they may be polaroids or nicols). C is either a glass container filled with some solution, or a sheet of some crystal. In the absence of C, let Q be "crossed" with respect to P. This means that if P transmits plane-polarized light having vibrations in some plane PP, then Q is set to transmit vibrations in a plane qq which is perpendicular to pp. So the net transmission is zero.

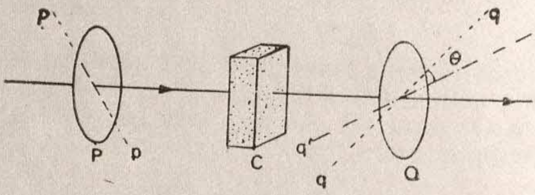


Fig. 17·1.

Optical rotation

Now let C be introduced. We note that with some samples the system now transmits light. But by rotating the analyzer Q by a certain angle θ the transmitted light can again be reduced to zero. This means that in passing through C the light remains plane polarized, but there is *rotation of the plane of polarization.*

This phenomenon is called "*optical activity*" or "*rotatory polarization*" or, more fully, "*rotation of the plane of polarization*".

Substances which show this phenomenon are called *optically active* substances. Those which rotate the plane of polarization clockwise are called *dextro-rotatory or right handed.* Those that rotate the plane of polarization anticlockwise (as viewed by the observer) are called *laevo-rotatory* or *left-handed.*

Several crystals show the phenomenon, the most important example being that of quartz, when a beam travels in it along the optic axis. Several organic substances in solution also show the phenomenon, the most important example being that of cane-sugar, which is right-handed.

17·2. Laws of Rotation of Plane of Polarization. The following laws have been observed experimentally.

(*i*) The angle of rotation (θ) of the plane of of polarization for a given medium (solid, liquid, gas or solution) for a given wavelength of light is proportional to the length (l) of medium traversed :

$$\theta \propto l$$

(*ii*) For vapours and solutions, the angle of rotation of plane polarization for a given path length is proportional to the concentration (c) of the vapour or solute :

$$\theta \propto c$$

This law, however, holds for small ranges of c only. For example, in potassium tartarate solution if we increase the concentration 8-fold (from 8 gm/100 cc) the rotation increases 9·1 fold.

(*iii*) If several media used separately cause rotations θ_1, θ_2, θ_3, etc., then if they are mixed together and are chemically non-reactive, the resultant rotation of the plane of polarization by the mixture is the *algebraic* sum of θ's.

$$\theta = \theta_1 + \theta_2 + \theta_3 + \cdots\cdots$$

This law does not rigorously hold in all cases.

Specific Rotation is defined differently for (*i*) solids, (*ii*) liquids, and (*iii*) solutions.

For solids, the defining equation is

$$\theta = \alpha\, l$$

where θ is the rotation measured in degrees, l is the path length measured in *millimeters*, and α is the specific rotation in units deg. mm^{-1}.

For liquids, the path length is measured in *decimeters* (10 cm.), so that specific rotation is given in units deg. dm^{-1}, eq. (17·1) being still applicable,

For solutions, the defining equation for α is

$$\theta = \alpha\, cl, \qquad \ldots(17{\cdot}2)$$

where c is concentration of the solute in gm per c.c. of the solution and l is path length measured in *decimeters* (10 cm). Hence units of α are deg. dm^{-1}. (gm/cc)$^{-1}$. Sometimes the concentration is given in gm of solute per gm of solution and symbol c is replaced by d in eq. (17·2). Then α has the units deg. dm^{-1}.

We give below the speific rotations of a few substances in solutions. Dextro and laevo-rotations are expressed by + and — sign respectively. The solvent is water, temp. 20°, $\lambda = 5893$ A°.

TABLE 17·1. *Specific Rotations of Some Materials*

Substance	*Specific rotation*
Cane Sugar	$+66{\cdot}5^\circ\ \mathrm{dm}^{-1}\ (\mathrm{g/cc})^{-1}$
Invert Sugar	$-19{\cdot}7$
d-glucose	$+52{\cdot}5$
α-fructose	$-11{\cdot}4$
Rochelle Salt	$+29{\cdot}7$
Nicotine	-77

Example 17·1. 20 gm of of cane-sugar is dissoved in water to make 50 cc of solution. A 20 cm length of this solution causes $+53^\circ\ 30'$ optical rotation. Calculate α.

Solution.

$$\theta=53{\cdot}5^\circ, \qquad c=20/50=0{\cdot}40\ \text{gm/cc} \qquad l=2\ \text{decimeters.}$$

$$\therefore\quad \alpha=\frac{\theta}{cl}=\frac{107}{2}\times\frac{1}{\cdot 40\times 2}=66{\cdot}9\ \frac{\text{deg}}{dm}\cdot\left(\frac{\text{gm}}{\text{cc}}\right)^{-1}$$ ⊙

Example 17·2. If 20 cm length of a certain solution causes righthanded rotation 38°, and 30 cm path of another solution causes left-handed rotation 24°, what optical rotation will be caused by 30 cm length of a mixture of the above solutions in volume ratio 1 : 2 ? (The solutions are not chemically reactive).

Solution. The 30 cm length of the mixture may by treated as 10 cm o the solution and 20 cm of the second (ratio 1 : 2).

$\therefore$ optical rotation by first is $38^\circ\times 10/20=19^\circ\ (+)$

,, ,, second ,, $24^\circ\times 20/30=16^\circ\ (-)$

$\therefore$ Total rotation $=19^\circ-16^\circ \qquad =30^\circ\ (+)$ ⊙

Example 17·3. Quartz has specific rotation 21·7° per mm for $\lambda=5893\ A$. A plane polarized beam is passed through a dextro-rotatory quartz plate of thickness 1·92 mm. Deduce the optical rotation.

Solution.

$$\theta=\alpha l=+21{\cdot}7\frac{\text{deg}}{\text{mm}}\times 1{\cdot}92\ \text{mm}=+41{\cdot}7\ \text{deg.}$$ ⊙

Rotatory Dispersion. The specific rotation α varies considerably with wavelength. Hence if a non-monochromatic plane polarized beam is passed through an optically active medium, lights of different wavelengths will have their planes of polarization rotated by different amounts. This phenomenon is called *rotatory dispersion.*

In general, light of shorter wavelength suffers larger rotation. In fact, for several substances α_λ varies *approximately* as $1/\lambda^2$, but this law is not quite rigorous (see Table 17·2).

TABLE 17·2 : *Variation of optical rotation of quartz with wavelength*

λ	α_λ	$\alpha_\lambda.\lambda^2$	λ	α_λ	$\alpha_\lambda.\lambda^2$
4358 A°	41·55°/mm	7191*	5893	21·72/mm	7543
4861	32·76	7745	6438	18·02	7298
5086	29·78	7709	6708	16·53	7439
5461	25·53	7614	7281	13·92	7379

17·3. Origin of Optical Activity. The fact that angle of rotation is proportional to path length l shows that the action of rotating the plane of polarization is not a surface effect but a body effect, *i.e.*, occurs inside the medium continuously. Further, the laws $\theta \propto$ concentration and $\theta = \Sigma\theta_1$ for a mixture show that it is a molecular phenomenon each molecule of the active substance playing its role *largely* independent of the presence or absence of other molecules. Substances which show optical activity in liquid state or in solutions must therefore have in their molecules some rotational symmetry. It is further interesting that most† optically active substances have both dextro and laevo-forms, *e.g.*, *d*-lacaic acid, *l*-lactic acid. The two have all physical and chemical properties identical, except that the optical rotations are of opposite sign and that they crystallize into hemihedral forms in opposite directions, like mirror images. It follows that the phenomenon of optical rotation *in organic liquids and solutions* is directly associated with structural asymmetry in the molecules. Fig. 17·2 shows a carbon atom *C* (tetravalent) surrounded by 4 *different* atoms *P*, *Q*, *R*, *S* in tetrahedral arrangement. On the left diagram *P*, *Q*, *R* are arranged clockwise (as seen along line *SC*), while on the right diagram they are arranged anticlockwise. Seen along *PC* we again find *Q*, *R*, *S* arranged in opposite order in the two cases, and so on. One structure is mirror image of the other. No rotation in space can make one structure coincide with the other. If this mirror-image asymmetry is the cause of optical rotation, it is apparent that a *tetravalent* atom, *e.g.*, carbon, sulphur, etc., attached to four *different* atoms or groups is an essential requirement for such asymmetry and hence for optical activity in liquids and solutions. This explanation has been widely found to hold true.

---

* Arbitrary units.

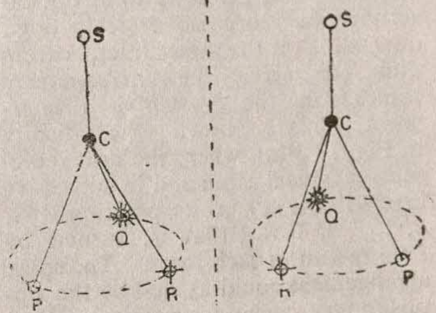
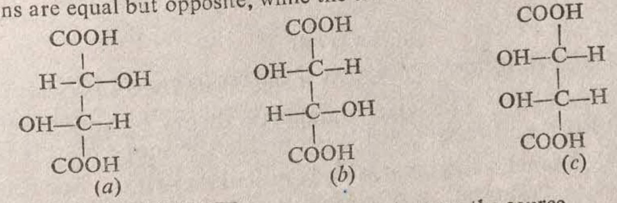


Fig. 17·2

Mirror-image symmetry

Tartaric acid presents an interesting illustration. Structural formulate of dextro (*d*—), laevo (*t*—) and meso tartaric acid are given below. Each of the two *C* atoms has four dissimilar groups (*H*, *OH*, *COOH* and the other *C*) attached to it, and there are two such *C* atoms in each molecule. In *d*-tartaric acid, both *C* atoms have *H*, *C.*, *OH*, *COOH* in one cyclic order ; in *l*-trataric acid again both *C* atoms have *H*,*C*, *OH*, *COOH* in one cyclic order which is opposite to that of the former ; in meso-tartaric acid the orders of *H*, *C*, *OH*, *COOH* in the two groups are opposite to one another. We find that in the first two the optical rotations are equal but opposite, while the third molecule (meso tartaric acid) is



$$\begin{array}{c} COOH \\ | \\ H-C-OH \\ | \\ OH-C-H \\ | \\ COOH \end{array}$$

(*a*)

$$\begin{array}{c} COOH \\ | \\ OH-C-H \\ | \\ H-C-OH \\ | \\ COOH \end{array}$$

(*b*)

$$\begin{array}{c} COOH \\ | \\ OH-C-H \\ | \\ OH-C-H \\ | \\ COOH \end{array}$$

(*c*)

Structural formulae of (*a*) *dextro-*, (*b*) *laevo-*, *and* (*c*) *meso-tartaric acid* not optically active. Thus, we may attribute some optical activity with *each group* within the molecule. In case (*a*) the activities of the two groups add up, so also in case (*b*), but in the third case (*c*) the activitives of the two groups, being opposite, cancel out.

---

† Cane-sugar is always dextro-rotatory, whatever the source.

Another situation for optical activity arises when the molecules have spiral-like arrangement in themselves, as schematically shown in Fig. 17·3 (*a*). The explanation of optical activity in inorganic crystals (*e.g.*, quartz), is on the same lines, except that the screw-like arrangement comes from the *crystalline arrangement*. This is shown schematically in Fig. 17·3 (*a*), where the dot represents a typical atom and in successive layers in the crystal its position is 1, 2, 3, 1, 2, 3......, that is, turned by $2\pi/3$ radian in each steps. The spiral arrangement could as well be the opposite way, causing opposite optical rotation.

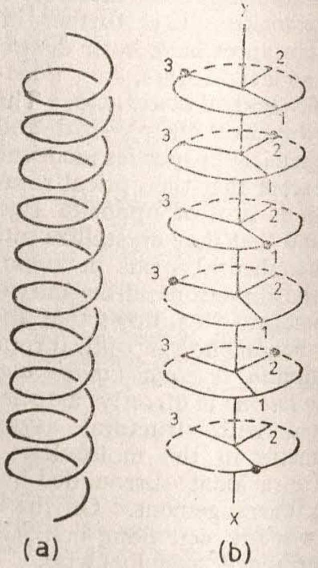


Fig. 17·3.

Spiral-like arrangements

Fig. 17·4 shows the spiral arrangement of silicon (Si) and oxygen (O) atoms in quartz in successive crystaline layers perpendicular to the axis *XX*.

Reuch gave an interesting demonstration of this theory. Mica is optically inactive. From a single cleaved thin sheet, several pieces were cut and then piled one over the other with their principal planes systematically turned in successive sheets. The pile showed optical rotation ; an opposite arrangement produced the opposite rotation.

It must be mentioned that these are *not theories*, but attempts to seek physical explanations and correlation of facts, on which a theory is to be built. The electromagnetic theory of optical activity is given by Born and others, but it is outside the scope of this book.

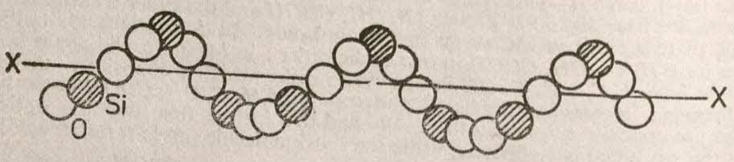


Fig. 17·4.

Spiral arrangement of Si and O atoms in quartz (SiO_2)

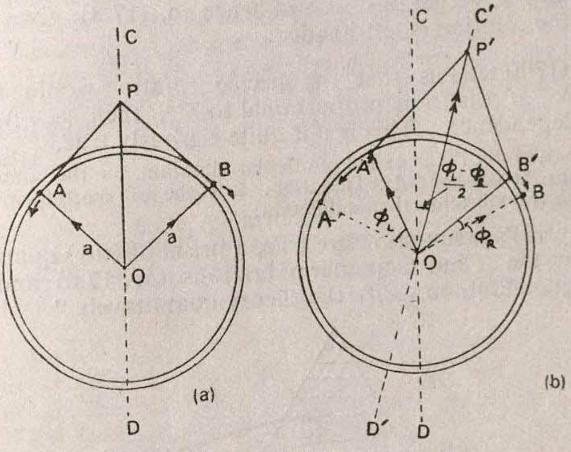
17·4. Fresnel's Explanation of Optical Activity. For rotation of the plane of polarization of a beam as the beam travels in some solutions or materials, Fresnel gave the following explanation :

(*a*) A plane-polarized vibration on falling on an optically active medium breaks into two circularly polarized beams of equa amplitude and *opposite* senses.

(*b*) The right circular and left-circular vibrations travel in the medium with different speeds V_R and V_L, so that a relative phase difference is created between them as they travel ahead.

(*c*) On emergence, the two circular vibrations recombine to give a plane-polarized vibration.

In Fig. 17·5 (*a*) consider two equal vectors *OA*, *OB* and their resultant *OP*. (For convnnience, *A* and *B* are shown on circles of slightly different radii). Obviously *P* lies on the bisector of *AOB* Now, if *OA* rotates anticlockwise and *OB* clockwise, with *equal* time periods *T*, their resultant will always lie along a fixed line *COD*, oscillating with amplitude $2a$, where $OA=a$. Converse of this is Fresnel's first assumption.



(a) Fig. 17·5. (b)

For Fresnel's explanation of Optical rotation

Now let the beam travel a length *l* in ihe medium. Then the righthanded vibration at emergence is *behind* the corresponding incident vibration by an angle

$$\phi_R=\Delta t.\frac{2\pi}{T}=\frac{l}{V_R}\times\frac{2\pi}{T} \qquad \text{...(17·4)}$$

The situation at emergence is shown in fig 17·5 [*b*], Here we show *OB′ behind OB* in its right-handed motion by angle ϕ_R. Similarly, the state of left-handed circular vibration is shown by *OA′ behind OA* in its left-handed motion by angle

$$\phi_L=\frac{l}{V_L}\times\frac{2\pi}{T} \qquad \text{...(17·5)}$$

In this figure we show $\phi_L > \phi_R$, *i.e.*, $V_L < V_R$.

The resulant of OA' and OB' is OP', which does *not lie* on COD but lies on another line $C'OD'$. From the geometry of the figure, the rotation is right-handed and amounts to an angle θ_R given by

$$\phi_R = COC' = \tfrac{1}{2}(\phi_L - \phi_R) = \tfrac{1}{2}\,\frac{2\pi l}{T}\left(\frac{1}{V_L} - \frac{1}{V_R}\right) \qquad \ldots(17{\cdot}6)$$

Multiplying the numerator and denominator on the right by c, the speed of light in vacuum, and remembering that

$$cT = \lambda, \qquad c/V_L, = \mu_L, \qquad c/V_R = \mu_R. \qquad \ldots(17{\cdot}7)$$

we get

$$\theta_R = \frac{\pi l}{\lambda}(\mu_L - \mu_R) \qquad \ldots(17{\cdot}8)$$

If $V_L > V_R$, we get $\mu_L < \mu_R$ and hence eq. (17·8) gives ϕ_R negative, *i.e.*, the rotation is left-handed.

Eq. (17·8) tells us that the specific rotation would then be $\pi(\mu_L - \mu_R)/\lambda$. If it is to be proportional to λ^{-2}, then $\mu_L - \mu_R$ should have λ^{-1} dependence. This is not quite rigrously true.

Note carefully T, the period of circular vibration, is the same for right-handed as for the left-handed polarization. It is the of speed forward propagation of the two vibration that is different.

If Fresnel's assumption are true, a prism of (say) quartz should separate out the R and L circular vibrations (fig. 17·6) and we can determine μ_R and μ_L to verify the theory quantitavely.*

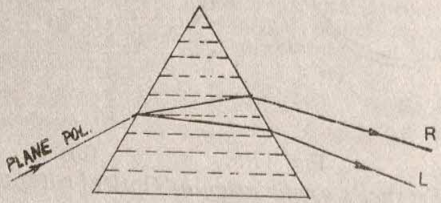


Fig. 17·6.

Right handed quartz $\mu_L > \mu_R$ *(schematic)*

Because $|\mu_R - \mu_L|$ is small, Fresnel used a train of alternate right-handed (R) and left-handed (L) quartz prisms, as shown in fig. 17·7, each having its optic axis normal to the end faces of the system. The left-handed and right-handed components separte more and more at successive interfaces. At emergence one can examine the two beams separately. They show circular vibrations in opposite senses.

---

* It may be mentioned that the difference $|\mu_R - \mu_L|$ is quite generally of a smaller order than the difference $|\mu_o - \mu_o|$. The angular spread shown in fig. 16·16 is realistic, while that in Fig. 17·6 is a large exaggeration.

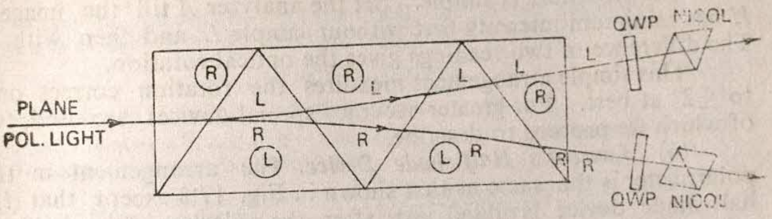


Fig. 17·7.

Train of R and L quartz prisms

17·5. **Polarimeters.** A device which measures the angle by which the plane of polarization of light is rotated by a given medium is called a *polarimeter*. By measuring this angle we can evaluate the specific rotation α, if concentration and path length are known. This measurement can be used to identity a substance. α may vary with wavelength or temperature or, more prominently, with various chemical substitutions in the molecule. Such measurements are widely used in chemistry to study the structure of molecules. Conversely, if α is known, then measurement of angle of rotation by a given path length leads to a knowledge of concentration of the active solute. Use is made of this principle in industry ; when used for evaluation of sugar content, the instrument is in fact named a *saccharimeter*.

(*a*) *Simple Polarimeter*. Fig. 17·8 depicts the schematic arrangement of a simple polarimeter. C is a collimator with a circular hole (H not slit) at the left. Light from source S is rendered parallel by the collimator. Polarizer P may be a polaroid or a nicol; it makes the beam plane polarized. Tube L contains the liquid or

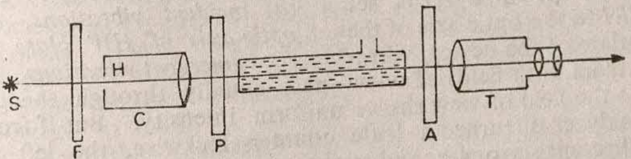


Fig. 17·8.

The simple polarimeter

solution under examination ; its two ends are covered by flat and parallel glass plates, whose separation is accurately known. This tube can be removed from position at will. Analyzer A may be a polaroid or a nicol ; it can be rotated about the axis of the system and the angle can be read on a circular scale. The transmitted light is observed through a telescope T,-which sees the image of the hole H. An optical filter F may be introduced anywhere in the path, the measured rotation will then be for the mean wavelength transmitted by this filter.

The procedure is simple. Set the analyzer A till the image of H has minimum intensity first without sample L and then with it. The difference of two readings gives the optical rotation.

This simple arrangement measures the rotation correct only to ±2° at best. For greater accuracy special devices are used, two of which we proceed to describe.

(*b*) *Laurent's Half-shade Device*. The arrangement in this polarimeter is the same as that shown in Fig. 17·8 except that (*i*) a half-shade device is added just after the polarizer P, and (*ii*) the telescope is focussed not on the hole H but on the half-shade device.

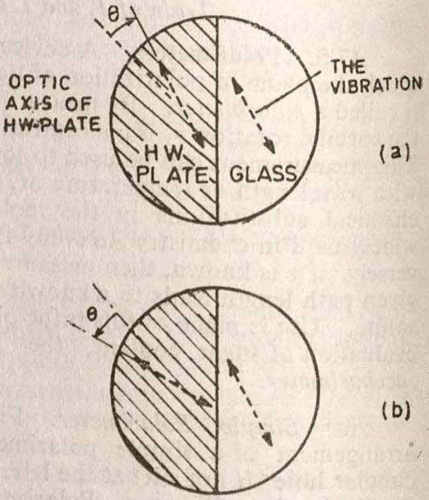


Fig. 17·9

The half-shade device of Laurent : (a) incident vibrations, and the optic axis of HW plate (b) the emergent vibrations

The half-shade device consists of two semi-circular plates one being of simple glass and other a half-wave plate for a particular wavelength (usually sodium D line 5893 *A*).

The principal plane of the polarizer is set to make a small angle θ with the optic axis of the HW plate Fig. 17·9 (a)]. The action of the *HW* plate is such that the emergent beam is again plane polarized but with vibrations inclined at engle θ on the otherside of its optic axis [Fig. 17·9(*b*)]. The glass plate, of course, passes the vibrations unchanged.

Now, when the principal plane of the analyzer is set *parallel* to the optic axis of the *HW* plate of the device, vibrations from both fields of view pass equally through the analyzer. Hence the field of view shows uniform intensity. But if from here the analyzer is turned a little counter-clockwise, the left field of view becomes stronger and right one weaker. If the analyzer is turned a little clockwise, the opposite thing happens. Thus a contrast of intensity in the two fields of view is created and hence setting at the uniform field position is quite accurate.

If the principal plane of the analyser is set *perpendicular* to the optic axis of the *HW* plate, there is again equality of intensity in the two fields of view. For small θ, this setting is more sharply sensitive than the parallel setting.

The half-shade device must be used with monochromatic light of that wavelength for which its *HW* plate is made.

Sensitvity of the half-shade device. In the true setting, the principal plane of the analyzer is $\perp^r$ to the bisector of the two vibrations, hence it makes an angle (90−θ) with each vibration.

If I_0 is the intensity of light from each half of the half-shade device, the intensity of light passed by the analyzer is, by law of Malus,

$$I=I_0 \cos^2 (90-\theta)=I_0 \sin^2 \theta \qquad ...(17{\cdot}9)$$

For a small change in angle $\Delta I=2\, I_0 \sin \theta \cos \theta .\ \Delta\theta$.

The intensity increases in one-half of the field of view, and decreases in the other half. Hence *contrast* may be expressed by

$$S \equiv \frac{\text{difference of intensities}}{\text{mean intensity}}=\frac{4\, I_0 \sin \theta \cos \theta \Delta\theta}{I_0 \sin^2 \theta}$$

or $S=4 \cot \theta .\ \Delta\theta$.

This shows that *for a given* $\Delta\theta$, *the contrast is larger when* θ *is smaller.* Of course, we cannot make θ too small, because the intensity is proportional to $\sin^2 \theta$ (eq. 17·9) and a certain minimum illumination is needed for the eye to judge the contrast.

(*c*) *The Biquartz Device.* It is a combination of two semi-circular discs of quartz, one left-handed (*LHQ*) and the other right-handed (*RHQ*), both cut with faces normal to the optic axis. Thickness of each is such that it rotates the plane of polarization for yellow light (~5900 *A*) by 90°. For longer wavelength the rotation is less, for shorter ones the rotation is more. Fig. 17·10 depicts what the *L* and *R* quartz plates do to incident plane-polarized white light.

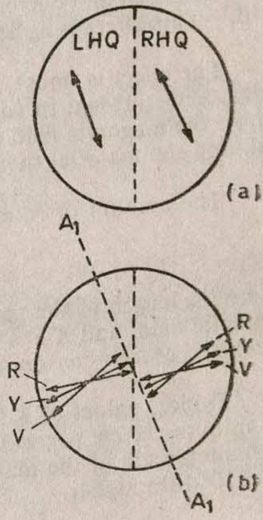


Fig. 17·10.
The biquartz device : (*a*) *incident vibrations* (*white*), (*b*) *emergent vibrations* (*dispersed*)

Now let an analyzer be set with its principal plane A_1A_2 parallel to the *original* incident vibrations. It can be seen that in this setting the emergent vibrations are equal for the two halves for each colour. Hence the two fields of view appear identical in overall colour. If the analyzer is turned a little clockwise, it more favourably passes red from the right half and violet from the left half. Thus, a sharp contrast of colours is created between the two halves. In a little anticlockwise turning the opposite contrast of colour occurs. Thus the analyzer can be set very accurately. The colour in both halves in this position is that arising from cutting off middle part of the visible spectrum and mixing the extreme colours. It is called '*tint of passage*' and any change in colour composition changes the tint very sharply. Hence the great accuracy of the device.*

---

* Another position of analyzer perpendicular to A_1A_1 gives equal tinge on the two halves. But that position is not sensitive. Why ?

In use with a polarimeter the biquartz device is fitted just after the polarizer (Fig. 17·8), and the observing telescope is focussed on the biquartz. The light used is not monochromatic but *white light.*

It is to be remembered that rotatory dispersion occurs not only in the quartz device but also in the sample medium which is under test. In one beam the dispersions due to quartz and the sample add up, in the other beam they subtract. The result is that no setting of the analyzer can truly bring equality of colour tinge in the two halves. Hence, if biquartz is to be used, the sample must be taken dilute or of small length so that total dispersion due to it is only a small fraction of the dispersion in the biquartz device.

17·6. Optical Rotation by Magnetic and Electric Fields. Faraday, in 1845, discovered that even isotropic media (*e.g.*, glass) cause rotation of the plane of polarization if they are subjected to a magnetic field parallel to the line of propagation. In principle the tube L in Fig. 17·8 may be replaced by the concerned dielectric (say a rod of glass) and current-carrying coils to generate the magnetic field. This effect is called the *Faraday effect.*

The effect is most pronounced in media of high refractive index, and is absent in vacuum. Hence, it is to be treated in two-steps : the magnetic field causing a suitable anisotropy in the material, and the material then causing the optical rotation.

The amount of rotation is given by

$$\theta = K B.\, l$$

where l is length of the path in the medium, B is strength of the magnetic field, and K is a constant (+*ve* or −*ve*) for the medium and is called *Verdet's constant.*

Typical values of K for organic substances fall in the range of 0·1 minute per *cm* per weber/m^2. Thus the effect is small. Yet it is characteristic of the material. For CS_2, flint glass etc., the values are$\sim 10^2$-fold higher.

Rotation of the plane of polarization is also observed under a transverse magnetic field. The effect is called *Cotton-Mouton effect,* and is an order of magnitude smaller than the Faraday effect.

Kerr discovered in 1875 that some transparent substances, which are otherwise isotropic, become doubly refracting under a strong electric field. As one result they rotate the plane of polarization of light passing through them. This effect is called *Kerr effect.*

Nitrobenzene shows the effect quite prominantly. Tube L in Fig. 17·8 is to be replaced by a chamber containing nitrobenzene and two parallel plates to create an electric field transverse to the beam of light. The chamber is called a *Kerr cell.* In § 7·5 we have alreadyshown one use of the Kerr effect.

In fact, when a plane polarized beam of light falls on the polished face of an electromagnet, the plane of polarization of the reflected light is not the same when the magnet is on as it is when the magnet is off. This was also discovered by Kerr and is referred to as Kerr effect by some authors and *Kerr Magneto optic effect* by others.

PROBLEMS

17.1. A solution of concentration 6·0 gm per 100 cm^3 used in a tube of length 33 cm causes 14° 30′ rotation in the plane of polarzation of light of $\lambda=5500$ *A*. Deduce the specific rotation. Also *estimate* the rotation it would cause for $\lambda=4500$ *A*.

17·2. State the reasons to conclude that optical rotation in liquids (*i*) is not a surface effect, (*ii*) has molecular origin. Discuss the origin of optical activity in liquids and solutions.

17·3. 20 cm path of a solution containing 10 gm of material *A* per 100 cm^3 causes optical rotation$+25°$, and 30 cm path of a solution containing 30 gm of material *B* per 100 cm^3 causes optical rotation$-15°$. Deduce the optical rotation due to 40 cm path of a solution containing 8 gm of *A* and 6 gm. of *B* in 50 cc of the solution. State any assumption made.

17·4. A material which is optically active but is not double refracting is put in the path of (*i*) an elliptically polarized light, (*ii*) a partially plane polarized light. Discuss the nature of the transmitted beam in each case.

17·5. An unpolarized beam of light is passed through a double image prism. Then one beam is passed through a quartz plate to rotate the plane of polarization by 90°, and this is then superposed with the other beam. Discuss if interference fringes would occur. How would things change if the starting beam itself were plane polarized ?

17·6. Outline Fresnel's explanation of optical rotation. Discuss (*i*) the dependence of rotation on λ, (*ii*) the experimental evidence in support of the assumptions.

17·7. In a Laurent half-shade device (Fig. 17·9), should the optic axis of the *HW* plate be parallel to the diameter ? Could the portion covered by the simple glass plate be left just vacant ? Where should the telescope of the polarimeter be focussed and why ?

17·8. In measuring concentration with a polarimeter, a sample tube of length $20{\cdot}0\pm0{\cdot}1$ cm is found to show an optical rotation $+32{\cdot}6\pm0{\cdot}1°$ with $\lambda=5893$ *A*. The specific rotation is known for this λ to be$+58{\cdot}2\pm01°$ per dm per gm/cc. Deduce the range of error in the concentration as obtained from these measurements.

17·9. The setting of a Laurent half-shade in a polarimeter was such that the optic axis of half-shade made 10° angle with the principal plane of the polarizer. Deduce the transmitted intenstiy and 'contrast for $\Delta\theta=0{\cdot}1°$'. How would these change if the setting is changed to 5° ?

17·10. For quartz the refractive indices for right-handed and left-handed circular vibrations of $\lambda=4000$ *A* are 1.55810 and 1·55821 respectively. Deduce the specific rotation for $\lambda=4000$ *A*. Also calculate the optical rotation caused by a plate 0·32 mm thick with its faces perpendicular to the optic axis.

17·11. Describe Faraday effect. What proves that the magnetic field does not act directly on the *e.m.* wave but only changes the configuration of the medium ?

17·12. Describe the Kerr cell, and any one situation where it is used as a very fast optical shutter.

18

Spectrographs and Various Kinds of Spectra

A spectrograph is an instrument to study the wavelengthwise distribution of light. The word 'light' in the spectroscopic sense is not limited to the visible range only. All the electromagnetic waves may be put under this name. However, our immediate interest is in the visible range and its neighbours—the ultraviolet and the infrared. We shall therefore, deal with instruments in these ranges only.

TABLE 18·1. *Different Regions of Electromagnetic Waves*

γ-rays	$\lambda \sim 10^{-11}$ cm
X-rays	$\lambda \sim 10^{-8}$ cm
Far ultra-violet	$\lambda \sim 2 \times 10^{-6}$ to 2×10^{-5} cm
Ultra-violet	$\lambda \sim 2 \times 10^{-5}$ to 4×10^{-5} cm
Visible	$\lambda \sim 4 \times 10^{-5}$ to 8×10^{-5} cm
Infra-red	$\lambda \sim 8 \times 10^{-5}$ to 3×10^{-3} cm
Far infra-red	$\lambda \sim 3 \times 10^{-3}$ to 3×10^{-2} cm
Micro-waves	$\lambda \sim 10^{\circ}$ cm
Radio waves	$\lambda \sim 10^{3}$ to 10^{5} cm.

The optical components of a glass prism *spectrometer* are familiar to the reader. If its eyepiece is removed and the cross-wires are replaced by a photographic plate, the instrument becomes a *spectrograph*. If, instead, a slit is placed in the position of the cross-hair and parallel to the spectral lines, the instrument becomes a *monochromator*. Quite often these terms are used indiscriminately, and where general reference is meant, we shall use the terms 'spectrograph'.

There are innumerable details in a modern spectrograph, particularly about the recording devices used. But our attention will be on the optics part *i.e.*, the formation of the spectrum.*

---

* Several interference devices for high resolution have been described in Chapter 12, and the concave diffraction grating in Chapter 14, with an eye on the wave theory.

18·1. Littrow Spectrographs. In Fig. 18·1, S is a slit, TRP is a total reflection prism, L is a collimating lens and P is a prism. The back face of the prism P is silvered, so that dispersed beam passes back through the prism and the same lens L acts for focussing the spectrum on the photographic plate PP. This is called a *Littrow arrangement*; the spectrum PP being effectively formed in the same plane as contains the slit S. The total reflection prism (TRP) is used to enable us to keep the slit S sideways.

In a Littrow spectrograph for use in the *ultra-violet*, TRP is replaced by a front-silvered plane mirror, L is a lens of fused quartz (silica) and P is a fused quartz prism. The photographic plate is then considerably inclined from the normal because the focal length of the silica lens varies considerably with wavelength in the ultra violet region. (See Fig. 18·4).

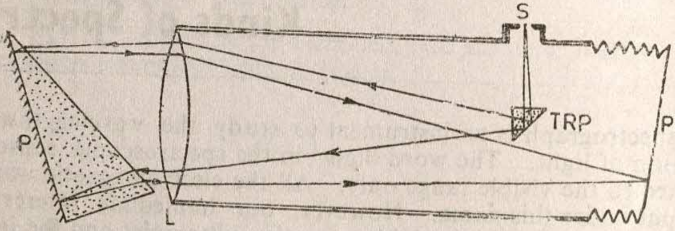


Fig. 18·1.

Littrows pectrograph with prism

We may also use a plane *reflection* grating in place of the prism P. By changing the inclination of the prism or the grating, different parts of spectra are brought in focus at the centre of plate PP in turns.

The Littrow arrangement is compact. Firstly, the same lens acts for collimation and focussing. Secondly, the prism or grating disperses twice. Thirdly, the slit and recording plate come near to make the setting handy. Another special feature is the absence of astigmatism in the image; the spectral image of a straight slit comes out straight. This is of special use in monochromators, where a straight exit slit is to be used. In § 18·3 we shall describe another Littrow spectrograph, where the Littrow mirror is kept separate from the prism.

Example 18·1. A plane diffraction grating with 5000 lines/cm is used in Littrow mounting with angle of incidence 20°, and focal length of collimator/camera lens 50 cm. Deduce for the second order spectrum (i) the wavelength falling at the centre of the field of view, (ii) linear dispersion in the spectrum (in Angstrom/mm).

Solution. (i) grating element $e=1/5000 \text{ cm}=2\times10^{-4}$ cm. In the Littrow mount $\theta\approx i=20^\circ$ (given). Hence, for second order,

$$2\lambda=e\,(\sin i+\sin\theta)=2\times10^{-4}\,(0{\cdot}34+0{\cdot}34)$$

$$\therefore\ \lambda=6{\cdot}8\times10^{-5}\text{ cm}=6800\ A.$$

(ii) Angular dispersion $d\theta/d\lambda$ for Littrow mount is given by

$$e \cos \theta \cdot d\theta = n d\lambda$$

∴ Linear dispersion $dx/d\lambda$ is given by

$$\frac{dx}{d\lambda} = f \cdot \frac{d\theta}{d\lambda} = f \, \frac{n}{e \cos \theta}$$

$$= \frac{50 \times 2}{2 \times 10^{-4} \times \cdot 94} \times 10^{-8} \text{ cm/angstrom}$$

$$\approx 5{\cdot}3 \times 10^{-2} \text{ mm}/A$$

$$\therefore \quad d\lambda/dx = 19 \; A/\text{mm}.$$

18·2. Constant Deviation Spectographs. The Littrow spectrograph of Fig. 18·1 is itself a constant deviation type, the 'constant deviation' in the case being 180°. In Fig. 18·2 we show a 90° *constant deviation prism KLMN*. As shown by the dotted lines *LN* and *OM*, the prism is equivalent to two 30°-right angled prisms *KLN*, *MON* and one 45°-right angled prism *LOM*. The 30° prisms cause dispersion, while the 45°-prism provides internal reflection to deflect the beam by 90°. Hence the total dispersion is equal to that of one 60°-prism. That light which is deviated by 90° passes through the prism in 'minimum deviation condition'. By rotating the prism with a suitable screw and drum-head, we can bring successive wavelengths in focus at the centre of the photographic plate or an exit slit. The drum is calibrated to read this wavelength directly.

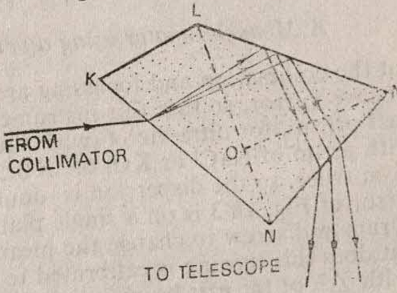


Fig. 18·2.
A 90° Constant Deviation Prism

Constant deviation spectrograpls of this design are made of glass (for visible) as well as quartz (for ultraviolet).

18·3. Monochromators. For intensity measurements in spectra we need to separate out a narrow range of the spectrum at a time. In priniciple this can be done in any spectrograph by replacing the photographic plate with a narrow slit whose length is parallel to the spectral lines. That would make the spectrograph a *monochromator*.

But two special considerations limit our actual choice. Firstly, for working at different λ values we should *not* have to move the exit slit, because other instruments (like photometer)) are linked with it. Hence, *only constant deviation spectrographs can be converted to monochromators*. Secondly, the spectral image should not be curved too much. For this reason arrangements with large astigmatic errors are not suitable for conversion to monochromators.

In Fig. 18·3 we show one monochromator. It uses a Littrow arrangement like Fig. 18·1 and has a constant deviation of 180°.

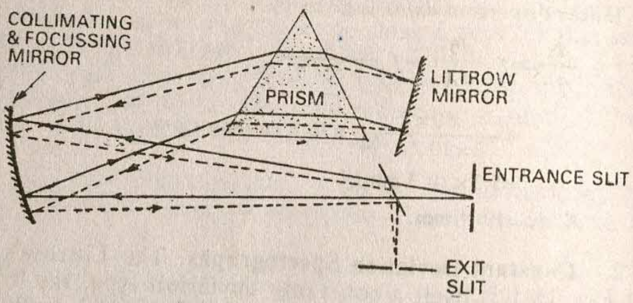


Fig. 18·3.

A Monochromator using a prism in a Littrow Mounting

But the collimation and focussing are here done by a front-silvered concave mirror, so that this instrument can be used for visible (with glass prism), for ultra-violet (with quartz prism), or for infra-red (with prism of NaCl or KCl or KBr, etc.) Also, this uses a full prism twice, so the dispersion is double. The prism plus Littrow mirror of Fig. 18·3 is on a single platform which may be turned by a drum and screw to change the mean λ of the exit beam. The drum that does this turning is calibrated to read the λ value. The spectral width △λ of the exit beam is broadly given by dividing the exit slit width *w* by the dispersion $dx/d\lambda$. But width of the entrance silt also is important.

18·4. Quartz Spectrograph. Glass is opaque for light of wavelength below∼4000 *A*, while quartz is transport down to a wavelength∼2000 *A*. Hence in the region 4000 to 2000 *A* quartz prism spectrographs are used.* Fig. 18·4 shows one.

Quartz prism is usually made of crystalline quartz, which shows double refraction. To avoid this the prism is so made that the optic axis is parallel to the base and in the principal section. Even in this arrangement, quartz shows optical rotation, which in Fresnel theory means breaking a beam into two left-handed and right-handed circularly polarized beams (Fig. 17·6). To avoid this effect a combination of two half-prisms is used, one of left-handed quartz and other of right-handed quartz. This is called *Cornu's prism.*

Quartz melted and re-solidfied is called *fused quartz or silica.* This is isotropic, *i.e.*, does not show double refraction or optical rotation. The lenses of the collimator and camera in a quartz spectrograph are made of fused quartz (silica).

* Quartz is also transparent for visible light. But the dispersion $d\mu/d\lambda$ for glass is much larger than that for quartz in the visible region. So prism spectrographs for the visible region use glass prisms only.

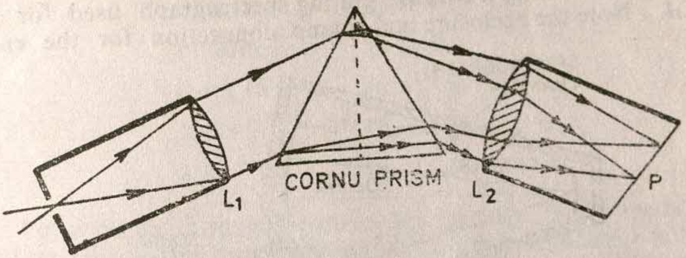


Fig. 18·4.

Quartz Spectrograph

Their focal length from 4000 *A* to 2000 *A varies widely*, but we cannot get achromatic combination in these lenses as we do for the visible region. Hence the difference of focussing distance for both collimating and camera lens has to be compensated by inclining the photographic plate (P) by a large margin (see Fig. 18·4).

High degree of homogeneity, required in prisms, is difficult to get in fused quartz. But in recent years technical advances have been made, so that prisms of fused quartz (silica) have replaced Cornu prisms in several instruments. Incidently, they are transparent down to 1850 *A*, which is an extra advantage.

18·5. Spectrographs for Far Ultra-Violet. The spectral range 4000 to 2000 *A* is often called the "*quartz region*" or "*near ultra-violet*". Below 2000 *A* is called *far ultra-violet*. Fluorite (CaF_2) is transparent in this region and spectrographs using flourite prism are therefore used.

There are other considerations. Since lenses of flourite are very expensive and have very large variation of focal length with varying wavelength, the collimation and focussing is usually done with front silvered concave mirrors. The arrangement is either as given in Fig. 18·3, or separate mirrors may be used for collimation and focussing as in Fig. 18·7.

Further, there is heavy absorption due to the atmosphere (mainly due to oxygen and water vapour) in the region of wavelength below$\sim$1850 *A*. Hence the whole system from the source to the spectrum must be enclosed in a chamber which is evacuated. For this reason this region of the spectrum is often called the *vacuum ultraviolet* and the spectrograph is called a *vacuum spectrograph*.

The fluorite (CaF_2) prism spectrograph is useful down to$\sim$1200 *A*. Below this we have not yet found a prism material sufficiently transparent. Then, we go over to *grating spectrographs*.*

---

* In fact grating spectrographs are useful for all regions of the spectrum from far ultra-violet to the infra-red. We have already described the concave grating and its mountings in detail in Chapter 14. Since no collimating or focussing device is separately needed with a concave grating it is really a very widely used spectroscopic equipment. However whenever prism spectrographs can be used, we choose to use them from intensity cosiderations. (see § 14·8).

Fig. 18·5 shows a concave grating spectrograph used for $\lambda <$ 2000 A. Note the enclosure and pump connection for the entire

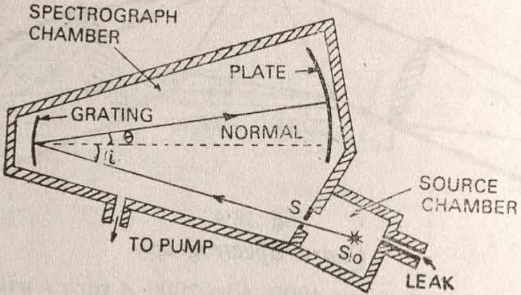


Fig. 18·5.

Concave grating spectrograph for vacuum ultra-violet region

spectrograph to keep vacuum. The source S_0 (say an arc) cannot work in high vacuum. Hence a separate chamber encloses it. It has a leak through which air (or any gas) keeps entering it. The slit S being narrow, continuous pumping keeps a lower pressure in the grating chamber than in the source chamber. Note also that with the usual grating element e about 16000 A, $(\sin i \pm \sin \theta)$ must be very small to record these lower wavelengths. For this reason the setting is such as to keep θ on the other side of i.

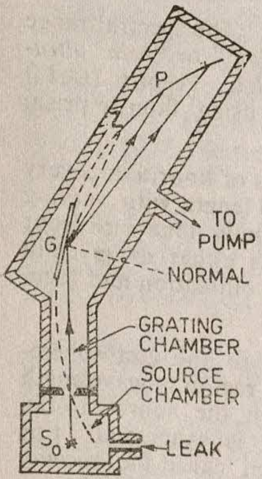


Fig. 18·6.

Concave grating in oblique incidence mount for very short wavelengths

The arrangement of Fig. (18·5) has been used down to $\sim$200 A. For still shorter wavelengths the reflecting power of even an aluminium concave grating becomes very small. Then oblique incidence is used (Fig. 18·6), for which reflectivity becomes far better. Another advantage of oblique incidence is higher dispersion, since $d\theta/d\lambda = n/(e \cos \theta)$ and becomes large on θ approaches $\pi/2$. In Fig. (18·6) the Rowland Circle is shown by the dotted line (passing through slit, grating and plate) and the zero-order is shown by Z.

18·6. Spectrographs for the Infra-red. Wavelengths in infra-red region are stated in the unit 'micron' (symbol μ), which stands for 1 micrometer $= 10^{-6}$ metre $= 10^{-4}$ cm $= 10{,}000$ A.

The useful *prism materials* are fluorite, rock-salt, KCl and KBr. The regions transmitted by them are as follows :

Glass* : 4000 A to $2{\cdot}5\mu$
Quartz : 2000 A to $3{\cdot}5\mu$
Fluorite : 3μ to 9μ
Rock salt : 8μ to 16μ
KCl : 15μ to 20μ
KBr : 18μ to 28μ

For wavelength range beyond $\sim 25\mu$ grating spectrometers are necessary. Below this λ they are useful, but loss of intensity prevents their general use. Blazed gratings (§ 14·11) are now available to get high intensity in a desired order.

Collimating and focussing in infra-red spectrographs is done invariably with concave mirrors. Various mountings are used, one of which we have already described in Fig. 18·3. In Fig. 18·7 we describe a Wadsworth mounting. In this, separate collimating and focussing mirrors are used ; the Wadsworth mirror just deviates the refracted beam in a convenient direction. Gradual rotation of the prism-mirror combination makes successive wavelengths reach the exit slit.

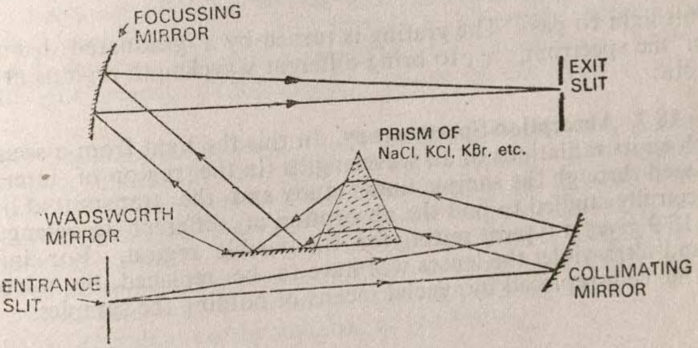


Fig. 18·7.

A typical prsim spectrograph for Infra-red. Recording part is placed behind the exit slit and is not shown

As detector for infra-red spectroscopic work a linear thermopile placed behind the exit slit is generally used. Its galvanometer shows deflections depending upon the intensity for the different wavelength regions brought at the exit slit.

Fig. 18·8 shows a typical grating spectrograph for infra-red. Notice that both the collimator and focussing concave mirrors are used at normal incidence to avoid astigmatism. Normal incidence working is achieved by the use of plane mirrors with an aperture to

---

* It is notable that even glass transmits upto 25000 A but its dispersion in this infared region is too small to give a useful spectrum.

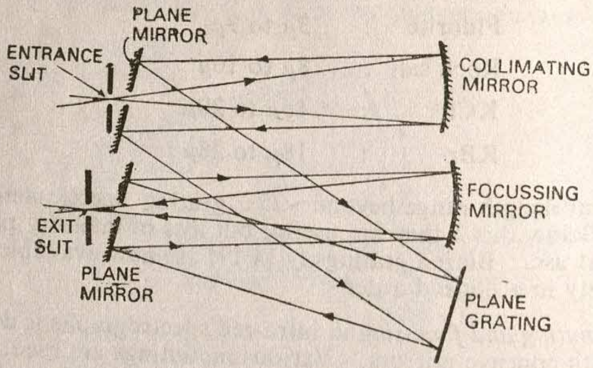


Fig. 18·8.

A grating spectrograph for infra-red

permit light to pass. The grating is turned by a graduated drum to 'scan' the spectrum, *i.e.*, to bring different wavelength regions at the exit slit.

18·7. Absorption Spectroscopy. In this the light from a source, which emits radiations of all wavelengths (in the region of interest) is passed through the sample under study and the transmitted light is spectrally studied to find the absorption at different wavelengths. Fig. 18·9 gives a typical setting for the visible region. For infra-red and ultra-violet the lenses will have to be replaced by mirrors and the 'cell' replaced by special means of holding the samples.

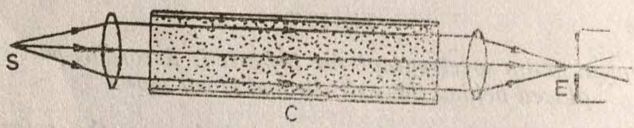


Fig. 18·9.

Typical setting for absorption spectrum : S is a filament source (giving continuous spectrum), C is either a cell containing the sample under test or a thin layer of the absorber placed directly in the path, E is entrance slit of a spectrograph.

In most absorption studies now the *double-beam principle* is used, depicted in Fig. 18·10. The incident beam is split at P_1 into two beams—one beam A passes through the absorption sample, while the other beam B passes through a variable intensity control R. The beams A' and B' are then to be balanced at each λ. The percentage cut-off of intensity needed at R to secure balance gives the percentage absorption by the sample in path A. The double-beam system makes the working independent of time variations in source intensity.

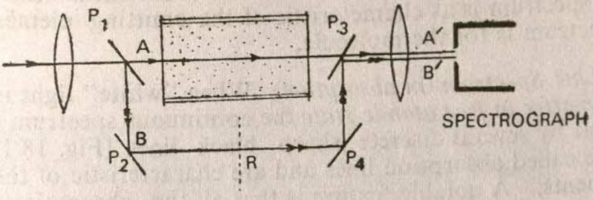


Fig. 18.10.

Principle of a double beam spectrophotometer. Plate P_1 splits the beams into A and B. One passes through the cell,the other passes through a variable rotating sector R

10·8. Some Typical Spectra. Spectra are studied in emission, absorption, reflectance, scattering etc. Each gives us information about the related syetem. But we will limit ourselves to spectra in emission and absorption. Fig. 18·11 is a schematic drawing, to be treated like the photographic negative—*i.e.*, black parts correspond to light present. A broad description of the spectra is as follows.

(*A*) *Continuous Spectrum*. Such a spectrum shows no discontinuity even under the best resolving spectrometer. It is emitted when *matter in bulk* is heated. A heated filament (as in bulbs), hot charcoal, hot iron, the photosphere of the Sun, etc., all give a continuous spectrum. The word "*white*" is usually employed to convey the sense of a continuous emission spectrum. This spectrum is largely characteristic of the temperature of the body and is only marginally dependent on the material of the body.

(*B*) *Band Spectrum* (*in emision*). Such a spectrum is emitted whenever *matter in molecular free state* is excited. Usually, each band has one end sharp [Fig. 18·11 (B)] and the other side shows tapering intensity.* CO_2, HCl, O_2, CO, AlO etc. when excited in vapour phase give band spectra. For each compound there could be several characteristic groups of bands falling in different spectral regions.

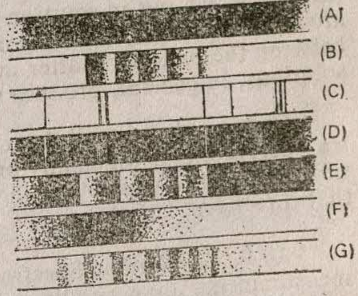


Fig. 18.11.
Some typical spectra

(*C*) *Line Spectrum* (*in emission*). Such a spectrum is emitted when

---

* Under high resolution instruments each band is seen as made up of several discrete lines—crowding up at the end which shows as 'head' and spreading out at the 'tail' end. In contrast, the continuous spectrum (A) does not show discreteness even under the highest resolution.

matter in free atomic state is excited. Discharge tubes containing Hg, Na, etc. and arcs using Cu, Fe, etc. as electrodes show line spectra. The line spectrum is as characteristic of the emitting element as the band spectrum is for the molecule.

(*D*) *Line Spectrum* (*in absorption*). When "white" light is passed through *matter in free atomic state* the continuous spectrum of white light is cut by several discrete sharp black lines [Fig. 18·11 (D)]. These are called absorption lines and are characteristic of the absorbing elements. A notable feature is that all the absorption lines of an element have their corresponding emission lines for the same element [compare Figs. (C) and (D)].*

(*E*) *Band spectrum* (*in absorption*). When white light is passed through matter in the free molecular state, the spectrum has discrete dark bands [Fig. 18·11 (E)]. Every absorption band of a compound has a corresponding emission band, but the detailed structures are not identical—the absorption band has simpler structure.

(*F*) *Continuous Absorption.* When "white" light is passed through matter in bulk, a certain very broad part of the spectrum may be absorbed (with no discreteness). Comparing Fig.(A) (white light) with (F) we find that in the latterc ase the left end of the spectrum is absent, and the right end is also somewhat weakened. A green or blue glass interposed between a white source and a spectrometer may produce such a spectrum.

(*G*) *Banded Spectrum.* (in absorption by solutions). A solution like potassium dichromate interposed between a white source and a spectrometer produces a spectrum which absorbs over a wide range, but the amount of absorption fluctuates alternately, so that bands are produced. We could interpret this in another way ; the corresponding compound in free state would absorb discrete bands (as in Fig. E), but in solution the bands are so broadened that there are no sharp edges, and no complete gaps of absorption appear in between the bands. Matter in molecular state (or ions) in dilute solutions, in general, produce such absorption.

This is only a broad description of the very vast variety of spectra observed, and which any proper theory of emission and absorption must try to explain. In the next chapter we make a beginning in that direction.

18·9. A Note on Spectroscopic Measurements. Spectroscopic measurements are generally expressed in terms of *wavelength.*

The corresponding *frequency* ν is given by

$$\nu = c/\lambda \qquad \ldots(18\cdot1)$$

---

* The reverse is not true. Thus the emission spectrum is always richer in the number of lines.

However, this becomes such a large number that we find it convenient to introduce "*reciprocal wavelength*" $\bar{\nu}$ or *wave number* defined by

$$\bar{\nu}=1/\lambda. \qquad \text{...(18·2)}$$

The unit for $\bar{\nu}$ is cm^{-1}, and a symbol K (named after Kayser) is introduced for it. Where there is no chance of confusion, reciprocal wavelength is often denoted by ν (without the bar on it).

One feature of spectroscopic measurements is the great degree of precision, often 1 part in 10^7. The corresponding requirement on any theories of atomic structure is that the theories have to explain details upto such degree of accuracy.

Example 18·2. Deduce $\Delta\nu$ values per angstrom difference of wavelength at $\lambda=2500$ *A*, and at $\lambda=5000$ *A*.

Solution $\bar{\nu}=1/\lambda \Rightarrow \Delta\bar{\nu}=-\Delta\lambda/\lambda^2$

For $\Delta\lambda=1\times10^{-8}$ cm at $\lambda=2500\times10^{-3}$ cm

$$\Delta\bar{\nu}=(1\times10^{-3})/(2{\cdot}5\times10^{-5})^2=-16\text{ cm}^{-1}$$

For $\Delta\lambda=1\times10^{-8}$ cm at $\lambda=5000\times10^{-8}$ cm

$$\Delta\bar{\nu}=-(1\times10^{-3})/(5\times10^{-5})^2=-4\text{ cm}^{-1}$$

[The minus sign only means that for greater λ the ν value is smaller.]

⊙

18·10. Raman Spectroscopy. When light of wavelength λ_0 passes through matter, some light is always scattered, and most of this scattered light has the same wavelength λ_0 as that of the incident light. Raman discovered in 1928 that there is also some *modified scattered light*, though it is quite weak. This is called *Raman scattering*, as distinct from coherent scattering. The modified spectral lines are called *Raman lines*, and the frequency shift

$$\Delta\nu_R=\nu_0-\nu_R=\frac{1}{\lambda_0}-\frac{1}{\lambda_R}$$

is called the *Raman Shift*. Raman lines occur generally on the longer wavelength side of λ_0, but with much lower intensity they also occur on the shorter wavelength side, in which case they are called *anti-Stokes lines*.

The important feature about Raman effect is that the *Raman shift $\Delta\nu_R$ is characteristic of the scattering molecules and is independent of the incident frequency as also of the direction of scattering.*

This makes Raman effect one of the most powerful tools in chemistry to investigate molecular structure. Previously, infra-red absorption spectra alone gave frequencies corresponding to molecular vibrations, rotations, etc. Now, the corresponding data are available through Raman effect in the visible and ultraviolet regions. What is more, some vibrations which do not show up in infra-red, show up in Raman effect and vice-versa. Further, the Raman spectra show characteristic polarizations, which give added information about the molecules.

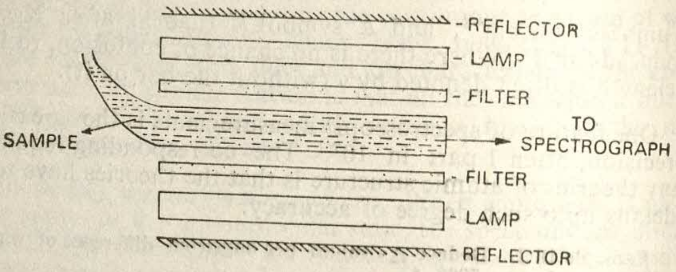


Fig. 18·12.

For Studying Raman Spectrum

Fig. 18·12 shows a typical conventional arrangement for observing a Raman spectrum. The sample is surrounded by a number of discharge lamps (typically Hg) with reflectors, and suitable filters are arranged in between to select one irradiating line or to filter out heat or both. The scattered radiation in the lateral direction is focussed on a spectrograph of high light gathering kind. A typical spectrum (with two Raman lines) is shown in fig. 18·13.

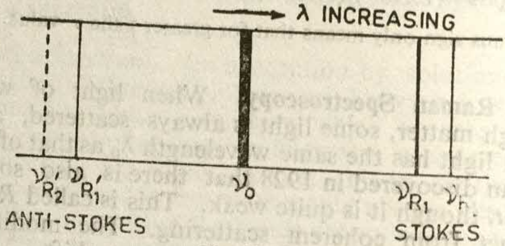


Fig. 18·13.

A Typical Raman Spectrum (Schematic)

With the advent of lasers (see Chapter 21) it has become possible to illuminate the sample with over a million-fold intensity. Hence Raman spectra can now be had very fast.

Example 18·3. Excited by 5460·7 $A°$ radiation a sample gives a Raman line at 5542·6 $A°$. Compute (i) the Raman frequency.(ii) position of the corresponding anti-Stokes line.

Solution $\nu_0=\dfrac{10^3}{5460{\cdot}7}=18312 \text{ cm}^{-1}$

$\nu_R=\dfrac{10^8}{5542{\cdot}6}=18042 \text{ cm}^{-1}$

$\therefore$ Raman frequency $\Delta\nu_R=\nu_0-\nu_R=270 \text{ cm}^{-1}$

anti-Stokes frequecy $\nu'_R=\nu_0+\Delta\nu_R=18312+270=18582 \text{ cm}^{-1}$

$\therefore$ anti-Stokes wavelength $\lambda'=\frac{10^8}{\nu'_R}=5381\cdot5$A

18·11. Luminescence Spectra. Several substances, after absorbing the incident light, emit light over a wide range of λ, characteristic of the sample. This phenomenon is called luminescence. If the emission is immediate (time-lapse$\sim10^{-8}\,s$), we call it *fluorescence*. In some cases the emission is delayed (time-lapse$\sim10^{-5}s$ to several hours), and then we call it *phosphorescence*. The important thing is that the emitted radiation is characteristic of the sample ; it does not depend on the exciting (incident) radiation, only the λ of exciting radiation has to be smaller than that of the characteristic emission.

The commonest use of fluorescence is in the mercury discharge tube used in homes. Fluorescent material on its inner walls absorbs the ultraviolet and other short λ radiations, and emits its own characteristic radiation, which falls in the visible range. Dyes are essentially fluorescent materials, and various dyes are available with their fluorescence peak λ falling in different regions. A yellow fluorescent paint in daylight shows larger intensity than the yellow part of solar radiation, because it converts a good part of the lower λ radiation also into its yellow emission.

Phosphorescence is a little more ticklish affair ; the substance gets excited by the incident light, but emits its characteristic light over a prolonged period. A study of the gradual decay of phosphorence with time under varying environment is of great interest.

PROBLEMS

18·1. Why is quartz spectrograph not suitable for wavelength (*i*) above 4000 *A*, (*ii*) below 2000 *A* ?

18·2. Linear dispersion in a spectrum may be increased by increasing the focal length of the camera. Why is this not used beyond reasonable limits ?

18·3. A plane reflection grating is used in a Littrow spectrograph. If the grating element is $1\cdot680\times10^{-4}$ cm, and the grating is set at $i=41°\ 30'$, deduce the wavelength at the centre of the spectrum in second order. Deduce also the linear dispersion, given f of lens $=1\cdot2$ metre.

18·4. Why should a monochromator be always based on a constant deviation principle ? Describe the optics of monochromator arrangements suitable for ultra-violet, visible and infra-red regions.

18·5. In the monochromator of fig. 18·3 replace the prism and Littrow mirror with a plane reflection grating. If $e=1\cdot84\times10^{-4}$ cm, deduce (*i*) the setting of the grating to bring the fourth order of $\lambda=5461$ *A* at the exit slit, (*ii*) the spectrum width passed by a 0·10 mm wide exit slit if focal length of the mirror is 60 cm, (*iii*) the rotation of the grating to change λ at the exit slit to 5487 *A*.

18·6. In a quartz spectrograph explain (*i*) the need of a Cornu prism, (*ii*) the large inclination of the photographic plate.

18·7. The dispersion in a particular spectrograph at 5000 *A* is 60 *A*/mm. Deduce (*i*) the range of spectrum passed by an exit slit 30 micron wide, (*ii*) the $\triangle\nu$ value for this band.

18·8. Classify the various kinds of emission spectra, and give examples of sources which emit them.

18·9. Describe a schematic arrangement for observing the absorption spectrum of a gaseous material.

18·10. One spectrograph resolves 2*A* at $\lambda=6000$ *A* ; another resolves 3*A* at $\lambda=4500$ *A*. Deduce the widths resolved in cm^{-1} for the two cases. Also compute $\lambda/\triangle\lambda$ and $\nu/\triangle\nu$ values.

18·11. An infra-red spectrograph resolves 0·31 micron at $\lambda=16$ micron. Deduce the $\triangle\nu$ resolved. Compare it with $\triangle\nu$ for an instrument of resolving power 200,000 at $\lambda=2500$ *A*.

18·12. With exciting radiation of $\lambda_0=4358{\cdot}3$ A°, a substance shows a Raman line at $\lambda=4624{\cdot}2$A°. Where will the corresponding anti-Stokes line fall ? Where will the Raman line fall for $\lambda_0=5460{\cdot}7$ A° ?

18·13. For a prism in minimum deviation position show that the angular dispersion is given by

$$\frac{d\theta}{d\lambda}=\frac{d\mu}{d\lambda}\cdot K$$

where *K* is nearly constant for a given setting.

18·14. State the advantages of a double-beam spectrophotometer over a single-beam one. What does the variable rotating sector do and why is it used ?

18·15. A grating of *e*=20000 A° is used to measure λ in the neighbourhood of 500 A°. Deduce the dispersion in the first order for angle of incidence (i) 0°, (ii) 75°0′.

18·16. Absorption spectroscopy is more widely used than emission spectroscopy. Explain why ?

18·17. In what way does Raman Spectroscopy help us in studying molecules ?

19

Origin of the Quantum Theory

A complete study of light involves two distinct aspects :

(i) the *distribution* of light energy in space when it passes across media, apertures, obstacles, etc.

(ii) the *exchange of energy* between light and matter when they interact in various ways.

So far in this book we have examined the first aspect. The propagation of light and its distribution in space could be quantitatively understood on the transverse wave theory of Fresnel. Maxwell's theory of electromagnetism fitted nicely into this, with the apparent advantage that the theory of matter and of light could now be unified.

However, just here the classical theories (*i.e*, Newtonian laws, thermodynamics and electromagnetic theory) gave results in conflict with experimental facts. *The interaction between light and matter could not be explained by the usual wave picture*, and around the year 1900 A.D. the house of physics was on fire.

The most important experimental facts which could not be explained on the basis of the existing theories at that time are :

(*a*) *The Black-body Spectrum.* The wavelength-wise distribution of light emitted by a hot enclosure.

(*b*) *The Photoelectric Effect.* The emission of photoelectrons from a given surface.

(*c*) *The Atomic Spectra.* The characteristic wavelengths of light emitted and absorbed by different elements.

(*d*) *The Specific Heats of Solids and Gases.* The low temperature specific heats of simple (elemental) solids and of diatomic gases.

The failure of classical physics forced physicists to search for a new framework of laws. That was the *background of the quantum theory*. In this Chapter we will introduce this background.

The crucial point that comes out is that the *exchange* in energy and momentum between light and matter occurs not in a continuous manner (as in waves), but in "*bundles*" or *quanta* of specified sizes. Since atoms and molecules are the ultimate emitters and absorbers of light, we necessarily have to introduce quantum concepts about the energy (and momentum) states of atoms too.

How are the energy states of atoms quantised ? A full answer to this question is beyond the scope of the present book. But we will attempt a partial answer.

In fact two distinctly different mathematical procedures (which have been shown to be equivalent in the final results) have been built up. One is called *Wave Mechanics*, and the other is called *Matrix Mechanics*. Usually the term *Quantum Mechanics* includes both.

Notice that any new theory must explain all the facts which were explained by the earlier theory, and *in addition* explain the hitherto unexplained facts. It may also *predict* some new results, which can be checked by further experiments. The quantum theory has this status. It leads to the same results as given by classical physics for gross observations ; but when it comes to details it leads to distinctly different results, in line with the experimental facts. As of date, all atomic, molecular and nuclear phenomena can be understood only through the quantum theory ; so also the properties of solids (like semiconductors), and new devices (like lasers, for instance).

We now turn to a survey of the experimental observations which lead to such interesting and exciting conclusions.

19·1. Black-body Radiation and Planck's Theory. A hollow enclosure maintained at a constant temperature T is called a "black-body".*

The radiation enclosed in its volume, and therefore the radiation emitted through a small hole in the wall, has got certain characteristics which are independent of the material of which the enclosure is made, and also independent of the shape and size of the enclosure. These characteristics depend upon the temperature alone.

The measurements presently of interest to us are the following :

(1) *Spectral Energy Density* $u\lambda$: the radiant energy per unit volume of the enclosure, which falls in unit range of wavelength at λ.

(2) *Spectral Emissivity* $e\lambda$: the energy radiated per unit area per unit time in unit range of wavelengths at λ.

It can be shown that for diffuse radiation $u\lambda$ and $e\lambda$ are related as

$$e\lambda=\frac{c}{4}u\lambda\,, \qquad \ldots(19\cdot1)$$

and therefore we may discuss any one of the quantities $e\lambda$ and $u\lambda$.

In fig. 19·1 we show an experimental plot of e_λ against λ. This is called a *spectral distribution curve* for black-body radiation. We repeat that this curve is found to depend on temperature alone, and is entirely independent of shape, size, or material of the enclosure. If the temperature is increased, there is increase of e_λ at all wavelengths, and the peak shifts to a smaller wavelength. But again for all isothermal enclosures of that temperature the same curve is *always* obtained.

---

* A small hole made in it will absorb 100% of the radiation falling on it. Hence, the name "black-body".

Among attempts to give a theoretical deduction of this distribution curve from classical considerations, two important land-marks may be mentioned. Wien proved from thermodynamics that

$$e_\lambda = \frac{1}{\lambda^5} F(\lambda T). \qquad \ldots(19{\cdot}2)$$

where F is some function of the product $\lambda\, T$. But it was shown that this function F could *not* be deduced from classical considerations. Rayleigh and Jeans (independently) used the method of statistical mechanics to show that

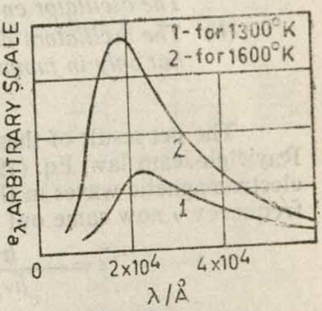


Fig. 19·1.
The black-body spectrum

$$u_\lambda\, d\lambda = \left(\frac{8\pi}{\lambda^4}\, d\lambda\right)(kT) \qquad \ldots(19{\cdot}3)$$

This has two parts : (*i*) $8\pi d\lambda/\lambda^4$ gave the number of standing wave modes per unit volume in the wavelength range λ to $\lambda + d\lambda$, and (*ii*) kT was taken as the mean energy per oscillation mode at temperature T (from classical law of equipartition of energy). Eq. (19·3) is called the *Rayleigh-Jeans law*.

Relation (19·3) contained no unknown parameter, and it fitted quantitatively in the longer wavelength parts of the experimental curves (Fig. 19·1) at all temperatures. But it failed totally on the shorter wavelength side. Going towards the ultra-violet (shorter λ) the experimental curves (Fig. 19·1) show a peak and then fall, while eq. (19·3) would indicate continuous rise only. This failure was often called "*the ultra-violet catastrophe*" of classical physics.

Wien's formula is called the *displacement law*. The reason is that if e vs, λ curve is known for any one temperature T_1, we can deduce from it the curve for any other temperature T_2. For this, measure the ordinate at any λ in the first curve, multiply it by $(T_2/T_1)^5$ and plot it at a new wavelength $(\lambda T_1/T_2)$. Thus the points in the first curve are displaced along λ axis and also along the $e\lambda$ axis by specific amounts. In particular, the peak of $e\lambda$ vs. λ curve occurs at wavelength λ_m given by

$$\lambda_m\, T = \text{constant} \qquad \ldots(19{\cdot}4)$$

This relation is used for measuring very high temperatures.

Planck in the year 1900 made a bold departure from the classical theory. He put forth two postulates :

(*a*) *An oscillator of frequency ν cannot have any general energy but only energies given by*

$$\varepsilon = nh\nu \qquad \ldots(19{\cdot}5)$$

where h is a universal constant (now called Planck's constant) and n is an integer (now called quantum number). The oscillator energy is thus quantized.

(*b*) *The oscillators do not radiate or absorb energy continuously but only in jumps or quanta of size hν*

$$\Delta\varepsilon = \Delta n . h\nu = h\nu \qquad \ldots(19{\cdot}5\,a)$$

The net result of these postulates was to change part (ii) in the Rayleigh-Jeans law, Eq. (19·3). The average energy per mode in the electromagnetic waves in the black-body chamber corresponding to frequency ν now came out as

$$\bar{\varepsilon} = \frac{h\nu}{e^{h\nu/kT} - 1} = \frac{hc/\lambda}{e^{hc/\lambda kT} - 1} \qquad \ldots(19{\cdot}6)^*$$

With this substitution for kT in eq. (19·3) we get what is called *Planck's formula* for black body spectrum :

$$u_\lambda d\lambda = \frac{8\pi}{\lambda^4} \cdot \frac{hc/\lambda}{e^{hc/\lambda kT} - 1} d\lambda \qquad \ldots(19{\cdot}7)$$

$$e_\lambda = \frac{2\pi c}{\lambda^4} \cdot \frac{hc/\lambda}{e^{hc/kT} - 1} \qquad \ldots(19{\cdot}7a)$$

For lorge λ the value of $\bar{\varepsilon}$ tends to kT and this formula coincides with Rayleigh-Jeans formula (19·3). At small λ the exponential term in the denominator prevents e_λ from rising. Thus the e_λ vs. λ curve has a maximum at some wavelength λ_m and falls off on both sides. This formula is found to give complete fit with the experimental observations at all wavelengths and for all temperatures.

Planck's postulate was the *origin of the quantum theory*. However, the experimental evidence was somewhat indirect, since it was (at that time) not clearly proved that Planck's assumption was the only way of explaining the iblack-body spectrum. The most direct evidence that energy exchanges can occur gn units of *hν* only came from the photoelectric effect, and the name *photon* was iiven to the quanta of energy *hν* by Einstein in 1905. The quantum theory got nto great strides only after that direct evidence.

Example 19·1. Deduce the average energy of Planck's oscillator corresponding to ν values such that $h\nu/kT = 0{\cdot}1,\ 1{\cdot}0,\ 3{\cdot}0,\ 10{\cdot}0$.

Solution. We write eq. (19·6) as

$$\bar{\varepsilon} = kT \frac{x}{e^x - 1}, \text{ where } x = \frac{h\nu}{kT}$$

For the four values of x tables give

$$e^x - 1 = 0{\cdot}105,\ 1{\cdot}718,\ 19{\cdot}1, \text{ and } 2{\cdot}2 \times 10^4$$

Hence the value of ε is given by

$$\varepsilon = kT\left(\frac{\cdot 1}{\cdot 105}, \frac{1}{1{\cdot}718}, \frac{3}{19{\cdot}1}, \frac{10}{2{\cdot}2 \times 10^4}\right)$$

---

* $\bar{\varepsilon} = \Sigma nh\nu \exp(-nh\nu/kT) \div \Sigma \exp(-nh\nu/kT)$ for $n = 0$ to ∞. The first part $nh\nu$ is Planck's quantum energy, the second part $\exp(-nh\nu/kT)$ is Boltzmann distribution law from thermodynamics.

$$\approx kT\,(1,\ 0{\cdot}58,\ 0{\cdot}16,\ 0{\cdot}00045)$$

Thus for $h\nu \ll kT$ we have $\varepsilon \approx kT$, while for $h\nu \gg kT$, ε tends to zero.

19·2. Photoelectric Effect and Einstein's Theory. In Fig. 19·2 *A* is a metal surface and *B* is a metal wire-gauge, both being enclosed in a glass bulb with high vacuum. The electric circuit is connected as shown. In the absence of light the circuit shows no current, but if light falls on the surface *A*, the circuit shows current if *B* is at positive potential. This is called *photoelectric current* ; the bulb enclosing *A* and *B* is called *photoelectric cell*. The effect is explained by assuming* that metals contain some electrons free to move *within* the metal and the light falling on surface *A* is able to eject electrons from the metal surface *A*.

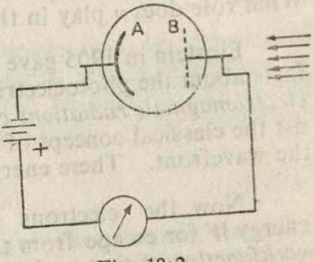


Fig. 19·2.
The photoelectric effect

If more light falls, more electrons are ejected, and therefore larger photoelectric current is obtained. This is actually observed. If electrode *A* is made positive, no current is observed even in the presence of light (ignoring minor details). This shows that light ejects only negatively charged particles from surface *A*. So far all good.

We now consider two more experimental facts :

(*i*) *No measurable time lag is observed between the falling of light and appearance of the photoelectric current, however weak the light may be.*

(*ii*) *If the frequency* ν *of incident light is decreased, then below a certain frequency* ν_0 *no photocurrent is obtained however large the intensity of light may be. Above this frequency* ν_0 *even the faintest light produces a photocurrent proportional to the intensity of light.*

As regards (*i*), simple calculation shows that a 10-watt source at 1 metre distance gives a flux $\sim 10^3$ erg cm^{-2} sec^{-1}. If this flux is distributed uniformly over the wavefront area, an average atom of cross-sectional area$\sim 10^{-15}$ cm^2 would receive energy at the rate of 10^{-12} erg sec^{-1}. If the electron requires 1 ev enegry† to come out of the metal, the atom would gather this energy in $\sim 1{\cdot}6$ sec. But experiments show no such time lag ; not even a millionth part of a second. In fact photoelectric effect is produced with far weaker light, meaning time lags of the order of days on the classical wave picture ! But no time lag is ever observed.

---

* The assumption was made to explain electrical conduction in metals.

† 1 $ev = 1{\cdot}6 \times 01^{-12}$ erg.

The existence of a limiting frequency ν_0 is also against the classical concept. A 100-watt source of $\nu<\nu_0$ would bring 100 times more energy per cm^2 per sec than a 1-watt source of $\nu>\nu_0$ placed at the same distance. How is it then that the 100-watt source produces no photocurrent while the 1-watt source does? What role does ν play in this game?

Einstein in 1905 gave the quantum theory to explain the above facts about the photoelectric effect. The essential postulate is that *electromagnetic radiations carry energy in bundles of unit* $h\nu$, as against the classical concept of continuous distribution of energy over the wavefront. These energy bundles are called *photons.*

Now, the electrons in a metal require a certain minimum energy W for escape from the surface. This energy is called the *workfunction* of that surface. If the photon energy $h\nu$ is greater than W then there is a *chance* that an electron may come out of the surface without time lag; but if $h\nu$ is smaller than W, there is no chance at all that an electron of the metal may get enough energy to come out of the surface.* The limiting frequency ν_0 (and corresponding wavelength λ_0) is thus connected with W through the relation.

$$h\nu_0 \equiv \frac{hc}{\lambda_0} = W \qquad \text{...(19·8)}$$

The Einstein postulate goes farther than Planck's postulate, since it applies quantization not only as emission and absorption steps in an oscillator but in the radiation itself. It explains the absence of time lag and also the existence of a limiting ν_0. But it also *predicts* some new things. If ν is greater than ν_0, the electron

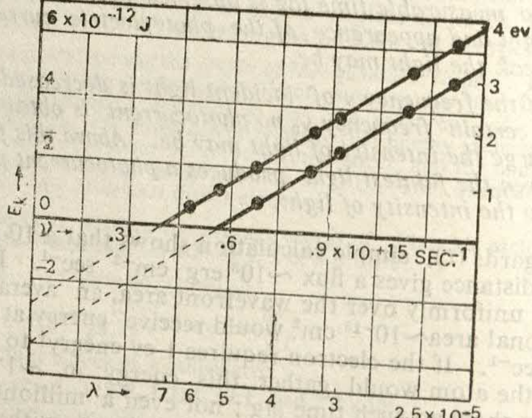


Fig. 19·3.

Plot of kinetic energy of photoelectrons against frequency of incident light ejected from the metal must come out with kinetic energy $h\nu - W'$ where W' is the energy needed to eject an electron in general from the given metal. The *minimum* value of W' has been called W, and hence the *maximum* energy of the ejected electron should be given by

---

* The chance of more photons than one imparting their energy to the same electron in the metal is negligibly small; see example 19·2.

$$E_k = h\nu - W = h(\nu - \nu_0) \qquad \ldots(19{\cdot}9)$$

The kinetic energy of the ejected electron may be determined by measuring its velocity v. But it is more simply measured by determining the *stopping potential* required to prevent the ejected electrons from reaching the collector electrode. If the *maximum* stopping potential is V_s we write

$$E_k = eV_s = h(\nu - \nu_0) \qquad \ldots(19{\cdot}9\ a)$$

Eq. (19·9) is called the Einstein equation of the photoelectric effect. It has been verified in full detail. In a series of careful experiments Millikan* obtained the stopping potential (hence E_k) for different frequencies ν of the incident light, and repeated this for several different surfaces of pure metals. He obtained experimental results like those shown in Fig. (19·3) for two different surfaces. The notable points are the following :

(*i*) For a given surface, the energy E_k of the fastest emitted electrons depends on frequency of the incident light ν alone and is totally independent of the intensity of light.

(*ii*) In each case E_k vs. ν graph comes out as a straight line.

(*iii*) The straight lines for different surfaces have the same slope.

From these experimental facts one could proceed in the reverse direction to deduce in three steps :

(*i*) $E_k = f(\nu)$, independent of intensity.

(*ii*) $E_k = h\nu - W$, where h is some slope.

(*iii*) h is a universal constant.

Of course, this slope is the Planck's constant of eq. (19·9). Its value came out to be the same as that of h used in eq. (19·5).

Example 19·2. A surface is placed 20 cm from a 50W source of $\lambda = 6 \times 10^{-5}$ cm. Estimate (i) the average time interval between the arrival of two photons on the same atom (sectional area$\sim 10^{-16}$ cm^2), and (ii) the mean time lag between irradiation and ejection of the first electron from the surface (area 2 cm^2, efficiency 40%).

Solution : Energy flux $= 50/4\pi\,(0{\cdot}20)^2 \approx 100\ \mathrm{Jm^{-2}\,s^{-1}}$

Photon energy $= hc/\lambda = 6{\cdot}6 \times 10^{-34} \times 3 \times 10^{+8}/6 \times 10^{-7} \approx 3 \times 10^{-19}$ J/photon

$\therefore$ Photon flux at 20 cm $= 100/(3 \times 10^{-19}) \approx 3 \times 10^{20}$ photon m^{-2}s^{-1}

Now, (i) photon flux on a given atom ≈ 3 photon s^{-1}

$\therefore$ mean $\Delta t \approx 0{\cdot}3$ s

---

* He was awarded the Nobel Prize in 1923 for this work.

(ii) ejection rate from 2 cm^2 area$=3\times10^{20}\times2\times10^{-4}\times\cdot4=2\cdot4\times10^{16}$ electrons s^{-1}

$$\therefore \text{ mean } \quad \Delta t\approx4\times10^{-15}\text{ s}$$

[*Note* : One conclusion is that the chances of two photons acting together in ejecting an electron are negligibly small. Another is that the time lag between irradiation and start of emission is too small to be observable] ⊙

Example 19·3. For a given metal surface the limiting wavelength for photoelectric emission is 5800*A*. Deduce the work function, maximum kinetic energy of electrons ejected by light of λ=4500*A*.

Solution. Photon energy $h\nu=\dfrac{6\cdot6\times10^{-27}\times3\times10^{10}}{4500\times10^{-8}}$

$$=4\cdot40\times10^{-12}\text{ erg.}$$

$$\text{Work function}=h\nu_0=\frac{6\cdot6\times10^{-27}\times3\times10^{10}}{5800\times10^{-8}}$$

$$=3\cdot41\times10^{-12}\text{ erg.}$$

The difference gives $\quad E_k=0\cdot99\times10^{-12}$ erg. ⊙

Example 19·4. Light of wavelength 5461*A* and intensity $0\cdot20\times10^2$ watt/cm^2 falls on a photoelectric cell of cathode area 2·5 cm^2. The efficiency is 15% Compute the saturation photocurrent.

Solution. Photons received *per second*

$$n=\frac{0\cdot20\times10^2\times2\cdot5}{6\cdot6\times10^{-27}\times3\times10^{10}\times1/(5461\times10^{-8})}=1\cdot37\times10^{13}\text{ s}^{-1}$$

Electrons ejected per second are 15% of this ; and each electron carries a charge $1\cdot6\times10^{-19}$ *coulomb*. Hence the saturation current is

$$i=1\cdot37\times10^{13}\times0\cdot15\times1\cdot6\times10^{-19}\text{ amp}$$
$$=0\cdot33\text{ microamp.}$$

⊙

Photovoltaic and Photoconductive Cells. Certain semiconductor sheets develop a potential difference across their faces when light falls on one face. This effect is called the *photovoltaic effect*, and a cell based on it is called a *photovoltaic cell*. The so-called solar cells are photovoltaic cells connected in a matrix to get voltage and power at useful levels.

Certain semiconductor materials show large increase in electrical conductivity when light falls on their surfaces. This effect is called *photoconductivity* and a cell based on it is called *photoconductive cell*. Usually a photoconductive cell is not sensitive in yellow and red regions, while a photovoltaic cell is.

In this context the cell described in Sec. 19·2 is called a *photoemissive cell*. We discussed that cell in detail because of its special importance in theory ; but incidentally it is also the one used most widely in practice.

Photomultiplier tubes. They use several intermediate electrodes between the photoemissive surface (cathode) and the final collector (anode). The photoelectrons produced at the cathode are accelerated through$\sim$100 volts, so that on striking the intermediate electrode (called dianode) each incident electron produces 2 (or more) electrons. In say, 10 such steps there is an electron multiplication$\sim2^{10}$ fold. The tube is thus really an *electronmultiplier tube*, and many authors use that name. Measurement of extremely low light intensities is possible with them; in fact photon counting systems are now in use.

19·3. Compton Effect. When light is scattered by any gas or liquid or solid, a major part of the scattered light has the same wavelength as that of the incident light. However, there is a part of scattered light with changed wavelength. In the visible and ultraviolet regions the most interesting modified scattered radiation is Raman scattering (see § 18·10). In the X-ray region the dominant modified scattering is that in which the change in wavelength ($\Delta\lambda = \lambda' - \lambda$) is *characteristic of the scattering angle, and independent of the scattering substance as well as the initial wavelength* λ. This effect is called *Compton effect*. It is further established that for every such scattered photon an electron is ejected from the target atom B (see Fig. 19·4). This is called a *recoil electron*. In order to explain his experimental measurements Compton in 1923 assumed that a photon of frequency ν must possess not only a definite *energy* $h\nu$, but also definite linear *momentum* $h\nu/c$. From this assumption and the elementary laws of collision he got the result

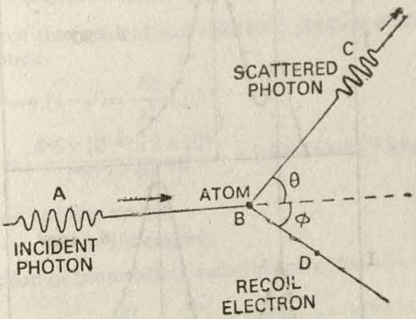


Fig. 19·4.*

Schematic representation of Compton experiment

$$\Delta\lambda \equiv \lambda' - \lambda = \frac{h}{m_0 c}(1 - \cos\theta) \qquad \ldots(19{\cdot}10)$$

Extensive experimental verification of this result gave support to his initial assumption about occurrence of momentum in quanta $h\nu/c$.

Proof of relation (19·10) uses relativistic expression for mass and kinetic energy of the recoil electron, because the speeds v are often close to the speed of light c. The expressions used are, with familiar notation,

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} \quad \text{and} \quad E_k = (m - m_0)\,c^2 \qquad \ldots(i)$$

Then using fig. 19·4 and conservation of energy and momentum we get

$$h\nu - h\nu' = (m - m_0)\,c^2 \qquad \ldots(ii)$$

$$\frac{h\nu}{c} = \frac{h\nu'}{c}\cos\theta + mv\cos\phi \qquad \ldots(iii)$$

and

$$0 = \frac{h\nu'}{c}\sin\theta - mv\sin\phi \qquad \ldots(iv)$$

---

* The photon representation used here is not valid, as will be seen in §23·4. It is only schematic.

Eliminating ϕ between (*iii*) and (*iv*)

$$m^2v^2c^2=h^2\,(\nu^2+\nu'^2-2\nu\nu'\cos\theta) \qquad ...(v)$$

From (*ii*) we get

$$m^2c^4=(h\nu-h\nu'+m_0c^2)^2$$

$$=h^2(\nu^2+\nu'^2-2\nu\nu')+2m_0c^2h(\nu-\nu')+m_0^2c^4$$

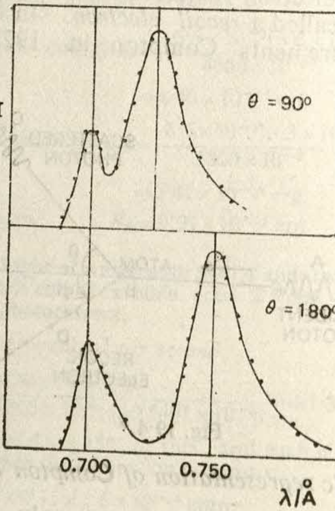


Fig. 19·5.*

Compton spectrum for θ=90° and θ=180°. The line at λ=0·700 A is the exciting line (λ_0). Vertical scale shows the intensity

Subtracting this from (v) gives

$$2h^2\nu\nu'\,(1-\cos\theta)-2m_0c^2h\,(\nu-\nu')=m_0^2c^4-m^2c^4\left(1-\frac{v^2}{c^2}\right)=0$$

$$\therefore\quad \frac{\nu-\nu'}{\nu\nu'}=\frac{h}{m_0c^2}(1-\cos\theta)$$

or

$$\lambda'-\lambda=\frac{h}{m_0c}(1-\cos\theta) \qquad ...(vi)$$

Substitution of constants gives

$$\lambda'-\lambda=0{\cdot}0242\,(1-\cos\theta)\ \text{angstrom.} \qquad ...(19{\cdot}11)$$

The quantity 0·0242 *A* is often called *Compton wavelength.* It is actually the *shift of wavelength* in Compton scattering at 90°.

---

* In Raman effect the ratio of modified to unmodified scattering is very small. But in Compton effect the ratio is large, and for large θ the modified (Compton shifted) component becomes stronger than the unmodified component.

In the deduction we have neglected the energy spent in detatching the electron from the atom. In fact we shall later see that an atom has several electrons with different binding energies. But an *X*-ray photon of $\lambda = 1A$ has energy $h\nu \approx 1{\cdot}2 \times 10^4$ *ev*, while the binding energies of the outermost electrons of an atom are only a few *ev*. Hence the binding energy may be neglected.

Example 19·5. In an experiment on Compton scattering the wavelength of incident radiation is 1·872 A. Calculate the wavelength of scattered radiation at θ=30°. Also calculate the velocity of the corresponding recoil electron.

Solution. By simple substitution in Eq. (19·11) for $\theta = 30°$

$$\triangle\lambda = {\cdot}0242\ (1 - {\cdot}8660) = {\cdot}00324A° \ ; \ \lambda' = 1{\cdot}872 + {\cdot}003 = 1{\cdot}875A.$$

The energy difference of the incident and scattered photons gives the KE of the ejected electron. Hence

$$\tfrac{1}{2}\, mv^2 = h\,(\nu - \nu') = \frac{hc}{\lambda^2}(\triangle\lambda)$$

$$= \frac{6{\cdot}6 \times 10^{-27} \times 3 \times 10^{10}}{(1{\cdot}872 \times 10^{-8})^2} \times ({\cdot}00324 \times 10^{-8})\text{erg}$$

Substituting for m and solving we get

$$v = 2 \times 10^8 \text{ cm/sec.}$$

Now, from conservation of momentum, remembering that $\nu'/\nu \approx 1$.

$$mv \cos\phi = \frac{h\nu}{c} - \frac{h\nu'}{c} \times {\cdot}866 \approx \frac{h\nu'}{c} \times 0{\cdot}134$$

$$mv \sin\phi = \frac{h\nu'}{c} \times 0{\cdot}500$$

$$\therefore\ \tan\phi = \frac{0{\cdot}5}{0{\cdot}134} = 3{\cdot}62, \qquad \Rightarrow \phi = 75°.$$

⊙

19·4. Series Relationships in Atomic Spectra. Each element, when excited in a gas phase, emits a spectrum containing sharp lines, which are characteristic for that element. If the characteristic wavelengths are $\lambda_1, \lambda_2, \ldots\ldots\ldots$ it is found that several of these can be grouped together and fitted into a single *empirical formula* with a variable number taking successive integral values. We call this group of wavelengths a *Spectral Series.*

The empirical relations are best expressed not in wavelength λ but in *reciprocal wavelengths* $\nu = 1/\lambda$, also called wave-numbers. The spectral lines falling in one series may be expressed as *differences* of terms of the following kind :

$$\frac{R}{(n-\alpha)^2}, \qquad \frac{R}{(m-\beta)^2} \text{ etc.,}$$

where R is a common constant for all kinds of terms and is called Rydberg constant ; α, β etc., are positive constants which vary for different kinds of terms ; and n, m etc., are positive integers taking values like $n = 3, 4, 5, \ldots\ldots$

As a particular example, one series in lithium spectrum is expressed by

$$\nu_n=\frac{109722}{(2-0{\cdot}040)^2}-\frac{109722}{(n-0{\cdot}405)^2},\ (n=3,4,5\ldots)\quad\ldots(19{\cdot}12)$$

Note that the first term is a *fixed term*, and the second one is a *running term* with $n=3, 4, 5\ldots\ldots$

Another series in lithium spectrum is given by

$$\nu_n=\frac{109722}{(2-0{\cdot}405)^2}-\frac{109722}{(n-0{\cdot}040)^2},\ (n=3,4,5\ldots)\quad\ldots(19{\cdot}13)$$

and yet another by

$$\nu_n=\frac{109722}{(2-0{\cdot}040)^2}-\frac{109722}{(n-0{\cdot}0026)^2},\ (n=3,4,5\ldots)\ldots(19{\cdot}14)$$

We find that these different series are not totally independent, but are mutually inter-related. The numerator 109722 cm^{-1}, common in all the terms, is the Rydberg constant. Further, in the first and third series the fixed term is identical, the running terms are different. Then the fixed term of the second series can be obtained from the running term of the first series by putting $n=2$. These relations are most generally expressed by the *Ritz Combination Principle* which says that any one of the running terms in one series can be used as the fixed term in another series to predict a new spectral series.

Example 19·6. One spectral series in an element is given by

$$\nu_n=\frac{109756}{(2-0{\cdot}083)^2}-\frac{109756}{(n-0{\cdot}643)^2},\qquad n=3,4,5,\ldots$$

Write down two more spectral series using Ritz principle, and evaluate ν for the first member of each series.

Solution. We predict one series by changing the integer in the first term denominator to 3. Thus

$$\nu_n'=\frac{109756}{(3-0{\cdot}083)^2}-\frac{109756}{(n-0{\cdot}643)^2},\qquad n=4,5,6\ldots.$$

We predict one more series by using the first running term of the given series as the fixed term, and changing the fixed term to running terms. Then

$$\nu_n''=\frac{109756}{(3-0{\cdot}643)^2}-\frac{109756}{(n-0{\cdot}083)^2},\qquad n=3,4,5,\ldots$$

For the new series thus predicted the ν values of the first term may be now calculated

$$\nu'=3{\cdot}17\times10^3\ \text{cm}^{-1}\ ;\ \nu''=6{\cdot}85\times10^3\ \text{cm}^{-1}.$$

⊙

Atomic hydrogen in a discharge tube gives various lines in its spectrum as shown fig. 19·6. Balmer discovered in 1885 that the lines in the visible spectrum of atomic hydrogen could be fitted in the formula

$$\lambda_n=K\,\frac{n^2}{n^2-4},\ \text{[Balmer Series]}\quad n=3,4,\ldots\quad\ldots(19{\cdot}15)$$

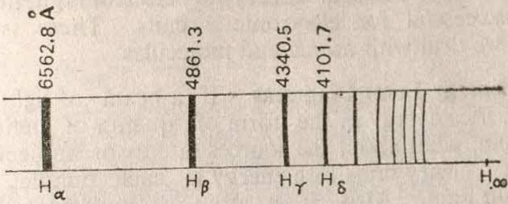


Fig. 19·6

The spectrum of atomic hydrogen

Balmer's formula takes a more convenient form if we use *wave-numbers*. Then

$$\bar{\nu}_n = \frac{4}{K}\left(\frac{1}{2^2}-\frac{1}{n^2}\right) = \frac{R}{2^2}-\frac{R}{n^2} \qquad n=3,\ 4,\ldots \qquad \ldots(19{\cdot}16)$$

where $R=4/K=109{,}678\ \text{cm}^{-1}$. The constant K is thus related with the *Rydberg constant.**

In this form Balmer formula has the same form as eq. (19·12) except that the denominators are simple integers. The formula, as before, tells us that the frequencies of various spectral lines in the series occur as *differences* of two terms.

We may now apply Ritz principle and predict that with any one of the running terms of eq. (19.16) as the fixed term, we should get other series :

$$\bar{\nu} = \frac{R}{1^2}-\frac{R}{n^2}, \qquad (n=2,\ 3\ldots) \qquad \text{Lyman Series} \qquad \ldots(19{\cdot}17a)$$

$$\bar{\nu} = \frac{R}{2^2}-\frac{R}{n^2}, \qquad (n=3,\ 4,\ldots) \qquad \text{Balmer Series} \qquad \ldots(19{\cdot}17b)$$

$$\bar{\nu} = \frac{R}{3^2}-\frac{R}{n^2}, \qquad (n=4,\ 5,\ldots) \qquad \text{Paschen Series} \qquad \ldots(19{\cdot}17c)$$

$$\bar{\nu} = \frac{R}{4^2}-\frac{R}{n^2}, \qquad (n=5,\ 6,\ldots) \qquad \text{Brackett Series} \qquad \ldots(19{\cdot}17d)$$

The names refer to the workers first discovering the series experimentally. Only the Balmer series falls in the visible region. It may be seen by comparion that the Lyman series falls in *UV* and Paschen and Brackett Series fall in *IR* region.

Many member lines of all these series have been observed and the experimental results agree with predictions very well. However, the basis for getting the frequencies $\bar{\nu}$ from *differences* of terms like $R/(n-\alpha^2)$ and the existence of a principle like that of Ritz could not be explained by the classical theory of electromagnetic radiation, which is so successful for electronic circuits. There is something special when we deal with atoms and molecules.

---

*Notice that this value of R for hydrogen is only slightly different from that for lithium, eq. 19·12 and that in Example 19·6.

19·5. Energy Level Diagrams. If a beam of light of a given wavelength carries energy in the form of quanta of definite energy and momentum, what about the sources (atoms or molecules) which emit the light ? They must lose energy in such bundles or quanta when they emit light. Also, if an atom or molecule absorbs light only for discrete frequencies, it must gain energy in steps only, *i e.*, in quanta, and not in a continuous manner.[†]

Take the example of sodium. In a discharge tube (or lamp) it emits most prominently the line 5893 *A* (we omit the detail that there are two lines 5890, 5896). This corresponds to quantum of energy $3·5\times10^{-12}$ erg. The emitted light carries energy as packets of this size. How are such packets emitted by the atoms of sodium ? Sodium vapour can also absorb light of $\lambda=5893$ *A* strongly. This again means that radiant energy which comes in packets of $\sim3·5\times10^{-12}$ erg. can alone be absorbed by the atoms of sodium ? Why is it that the atoms of sodium can emit and absorb packets of a particular energy only ?

For every element there is a discrete emission spectrum, *i.e.*, the emitted light consists of a few sharp frequencies, which are the same for the given element in all circumstances* and at all times, For every element there is also a discrete absorption spectrum, *i.e.*, out of all possible radiations which may fall on it, the element absorbs light of only a few sharp frequencies, which are the same for the given element in all circumstances* and at all times. Why ?

The questions lead us to assume that the atom of a given element can exist in several *discrete energy states* only. This means that the atom can either be in state *A* or in state *B* or in state *C*,..... but in no intermediate energy states. In Fig. 19·7, which is called an *energy level diagram*, the vertical gaps indicate the differences of energy[†]. It means that in a large assembly of these atoms some will be in energy state *A*, some in state *B*, some in state *C*..., but no atoms can acquire any intermediate energy state.

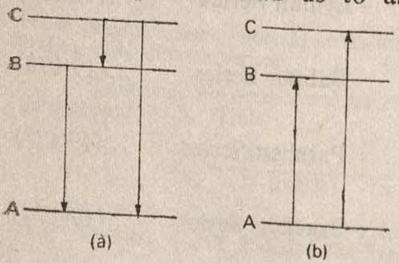


Fig. 19·7

Energy level diagram and trannsitions for (a) emission, (b) absorptioon

Now we further assume that when the atom changes from state *B* to state *A*, then the diference of energy E_B-E_A is given out as a photon (bundle of radiant energy), whose frequency ν_{AB} is given by the rule

---

[†] Similar statements must apply for *momentum* also.

*Minor variations of detail neglected.

[†] The horizontal lengths in Fig. 19·7 have no significance.

$$h\nu_{AB}=E_B-E_A \qquad \text{...(19·18)}$$

One atom passing from energy state B to A gives out one photon of frequency ν_{AB}. Thus a transition from state B to state A always produces light of the definite frequency ν_{AB}. More light of a given frequency means more atoms undergoing transition from state B to A.

We can similarly think of emission of photons of frequency $\nu_{BC}=(E_C-E_B)/h$, $\nu_{AC}=(E_C-E_A)/h$, etc. The transitions are shown by downward arrows in Fig. 19·7 (*a*).

It is also possible that an atom may pass on from a lower energy state to a higher energy state, say from A to B. For this it will need energy E_B-E_A to be supplied. It can be supplied in many ways, but suppose we consider only the case of supply through absorption of light falling on the atom. If its frequency is such that $h\nu$ is neither equal to E_B-E_A nor equal to E_C-E_A, then the photons cannot be absorbed, because the atom is not allowed to have intermediate energy values, and each photon is a single bundle of energy which cannot be taken in parts. So only light of such frequency for which $h\nu$ matches the energy gap E_B-E_A or E_C-E_A can be absorbed by the atom.

Thus we have a framework (a model) in which each ν arises due to a *difference* of two terms, and Ritz combination principle can be explained. We have to assume discrete energy levels and Eq. (19·18). Why do the atoms have discrete energy levels ? We shall see this later. But we at least realise at this stage that atoms existing in quantized states of energy and radiant energy existing in the form of quanta of energy are two interlinked concepts.

19.6. The Franck and Hertz Experiments. Direct evidence for quantized energy states in atoms comes from experiments of the kind first reported in 1914 by Franck and Hertz, and extended

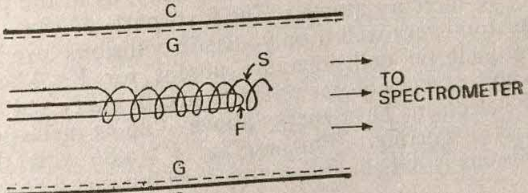


Fig. 19·8.

The Franck and Hertz experiment.

with different modifications by others. The principle is shown in Fig. 19·8. F is a tungsten filament heated (electrically) so as to release electrons. S is a spiral which is kept at a potential $+V$ with respect to F, and C is a metal cylinder at the same potential as S. The electrons from F gain kinetic energy eV as they travel from F to S, and should reach the cylinder C with the same energy, but for the grid G.

The whole arrangement is enclosed in a chamber, and pressure of the gas is so adjusted that in distance F to S the electrons have little chance of collision with atoms of the gas, but in distance S to C there may be *many collisions*. Thus, electrons of kinetic energy eV suffer many collisions with atoms of the gas in the tube before they reach C across the field-free cylindrical space between S and C. A spectrometer looks for any light emitted by the system.

The arrival of electrons at C is recorded by an electrometer (not shown). To find out the energy with which the electrons reach C, a grid G is put very close to C. It is given a negative potential V_G relative to C, and by finding that value of V_G which is just able to stop the electrons from reaching C we know that the electrons have energy eV_G when they arrive near C.

We describe the results with mercury vapour in the chamber. As V is increased from zero, we find that at first V_G also has to be increased right upto V to stop the electrons reaching C. This means that *in spite of collisions* with mercury atoms in the space between S and C, the electrons reach C without any noticeable loss of energy. The spectrometer shows no light. When V exceeds 4·86 volts, even a very small value of V_G, always equal to $V-4{\cdot}86$, is able to cut off the current almost completely. Also, the spectrometer, which showed no light so far, shows just at this stage a spectral line at $\lambda=2537A$.

Now, collisions between electrons and atoms can be of two kinds—*elastic* and *inelastic*. In an elastic collision there is no change in the *internal* energy of the atom. By applying conservation of kinetic energy and momentum one can show that for elastic collisions the energy loss per collision should be extremely small if the mass ratio of the colliding particles is far from unity, as in the present case of electrons and mercury atoms. The first part of the observation is thus understood very well if only elastic collisions are present. It indicates that inelastic collisions are absent for $V<4{\cdot}86$ volt. In other words the mercury atoms refuse to accept enery $<4{\cdot}86$ *ev*. But just when the electrons have energy above 4·86 *ev* inelastic collisions suddenly start occurring. Further, for $V>4{\cdot}86$ volt the loss in inelastic collisions is not the full energy V *ev* but just 4·86 *ev*.

We are therefore led to the conclusion that the atoms of mercury do not have continuous internal energy states. The first higher energy state is 4·86 *ev* above the lowest state. When the impinging electrons have energy exceeding 4·86 *ev* the atoms *may* accept this energy and pass on to the higher energy state, leaving the electrons with so much smaller energy.

The spectroscopic observation is now easily understood. For $V<4{\cdot}86$ volt no atoms reach the excited state (like B, Fig. 19·7) hence, no emission. For $V>4{\cdot}86$ volt, some atoms reach an excited state 4·86 *ev* above the lowest. When these atoms return to the lowest state (called ground state) energy 4·86 *ev* is emitted as light. The frequency of this light should be given by the rule $E=h\nu$. We check up :

$$E=4{\cdot}86\ ev=4{\cdot}86\times1{\cdot}6\times10^{-12}\ \text{erg}=7{\cdot}78\times10^{-12}\ \text{erg}.$$

$$h\nu=\frac{hc}{2537\times10^{-8}}=\frac{6{\cdot}6\times10^{27}\times3\times10^{10}}{2{\cdot}537\times10^{-5}}=7{\cdot}78\times10^{-12}\ \text{erg}.$$

As the voltage on S (Fig. 19·8) is further increased, the electrons reaching C have energy either V *ev* or $(V-4{\cdot}86)$ *ev*. But as the accelerating voltage V exceeds the value 6·67 volt, we find that a good proportion of the electrons reaching the collector C have energy $(V-6{\cdot}67)$ *ev* only. This shows that mercury atoms can be excited to one more energy state 6·67 *ev* above the ground state. Proceeding thus, we find other energy levels. Ultimately when V exceeds 10·4 volt we find that the electrons reaching the collector C have *any energy* smaller than $(V-10{\cdot}4)$ *ev*. It means that above 10·4 *ev* the energy levels for *Hg* atoms become continuous. We observe experimentally that for $E>10{\cdot}4$ *ev* the mercury atoms break into an electron and a positive ion each, *i.e.*, the atom is ionised.

We express the *experimental* results of this investigation on gaseous mercury by the energy level diagram of Fig. 19·9 (*a*). Similar *experimental* results for *Cs*, *Na* and atomic hydrogen are also shown there.

A related experimental result is that the spectrum shows more and more lines as the energy of electrons reaches each of the discrete energy values, the measured wavelengths always satisfying the relation

$$h\nu=\text{difference of energy}.$$

For example, from Fig. 19·9 (*a*) we expect that when the accelerating voltage V of electrons is $8{\cdot}84>V>6{\cdot}67$ only two energy states (4·86 and 6·67 *ev*) are reached by the mercury atoms, and hence the observed spectral lines should be at frequencies corresponding to $h\nu=(6{\cdot}67-0)$, $(4{\cdot}86-0)$, and $(6{\cdot}67-4{\cdot}86)$ *ev*. But as soon as V exceeds 8·84 volt three *additional* spectral lines should appear at $h\nu=(8{\cdot}84-0)$, $(8{\cdot}84-4{\cdot}86)$, and $(8{\cdot}84-6{\cdot}67)$ *ev*. This is exactly what is observed, *i.e.*, we have controlled excitation of the spectrum in steps.

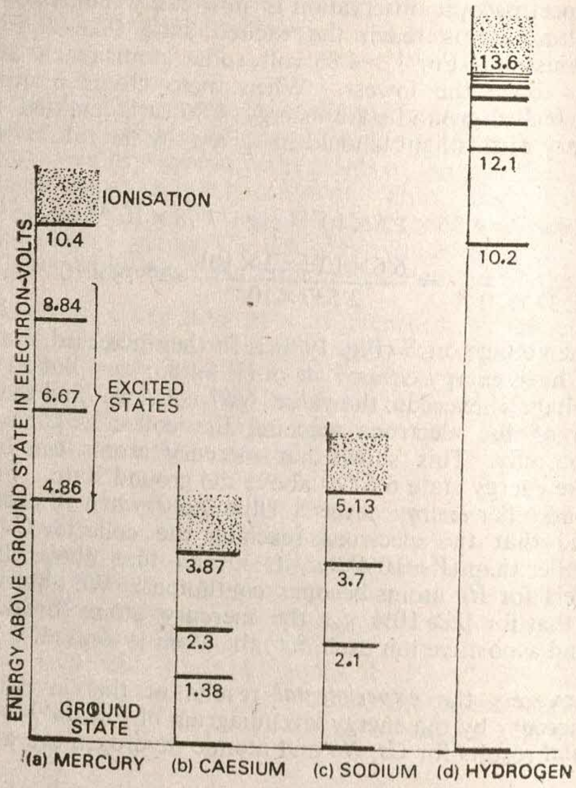


Fig. 19·9.

Energy states of some typical atoms. Only a few states are shown in each case

Conversely, if white light is passed through mercury vapour, then only those portions of the spectrum are absorbed where $h\nu$ values correspond to 4·46, 6·67, 8·84, and >10·4 *ev*.

Thus, Franck and Hertz experiments firmly establish that—

(*a*) atoms have discrete energy states.
(*b*) the energy states are different for atoms of different elements.
(*c*) ordinarily almost all atoms are in the ground state.* Emission of light takes place only if the atoms can be raised to one of the excited states.
(*d*) the frequency of emitted light follows the rule $h\nu = \Delta E$.

---

*The fact that absorptions corresponding to (6·67—4·86), (8·84—6·67), (8·84—4·86) etc. do *not* occur means that ordinarily there are almost no atoms in the states above the ground state.

Example 19·7. With Franck-Hertz type of experiment on sodium vapour the first spectral line to appear is the D-line, $\lambda=5.89\times10^{-5}$cm. Deduce the first excitation potential of sodium.

Solution. If excitation potential is V *volt*, excitation energy is eV joule. where $e=1{\cdot}6\times10^{-19}$ coul. Equating this with $h\nu$.

$$1{\cdot}6\times10^{-19}\ V=h\nu=\frac{hc}{\lambda}=\frac{6{\cdot}62\times10^{-34}\times3\times10^{8}}{5{\cdot}89\times10^{-7}}$$

$$V=\frac{6{\cdot}62\times3}{5.89\times1{\cdot}6}=2{\cdot}1\text{ volt.}$$

Example 19·8. Sodium has first two excited states at 2·1 and 3·7 eV. In a Franck-Hertz experiment electrons of energy 4·7 eV are fired in sodium gas. Deduce the possible energy values of the electrons received at the collector.

Solution. In collisions the electrons may lose energy 0, or 2·1 eV, or 3·7 eV or 2×2·1 eV or (2·1+37) eV, or (2×3·7) eV...Out of these the values below 4.7 eV are 0, 2·1, 3·7, and 4·2 eV. Hence the possible energies of the received electrons are 4·7 eV *minus* these values. Thus

$$E=4{\cdot}7,\ 2{\cdot}6,\ 1{\cdot}0,\ 0{\cdot}5\ eV.$$

19·7. The Specific Heats of Simple Solids. The specific heat *per gm-atom* of many elements in the solid state ranges from 5·38 to 6·93 cal (gm-atom)$^{-1}$ (deg)$^{-1}$ with an average of 6·15. This experimental fact is known as Dulong and Petit's law : *atomic specific heats of simple (elementary) solids are nearly equal.*

However, there are striking exceptions like diamond and silicon showing C_v much lower than 6 cal (gm-atom)$^{-1}$ (deg)$^{-1}$. Also, low temperature studies show that C_v of even those substances which follow Dulong and Petit's law at ordinary temperatures shows a large decrease as the temperature is lowered, tending to zero at $T=0$. Fig. 19·10 shows the typical results.

Now, in classical physics the average energy of an oscillator comes to *kT for all oscillators*. Hence if the solid is treated as made up of N oscillators (N=Avogadro number), each with 3 degrees of oscillatory freedom, the energy per gm-atom is

$$E=3NkT+E_b\ (N=\text{Avogadro number})$$

where E_b is the binding energy in going from free atoms to the soild state. Now, since E_b is largely independent of T,

$$C_v=\frac{\partial E}{\partial T}=3Nk=3R=6\text{ cal deg}^{-1}\text{ mole}^{-1}$$

This is a simple and elegant explanation of the Dulong and Petit's law. But the theory totally fails to explain the exceptions (*e.g.*. C and *Si*) and also the variation of C_v with temperature. Classical theory leads to $C_v=3R$ for all elemental solids at all temperatures.

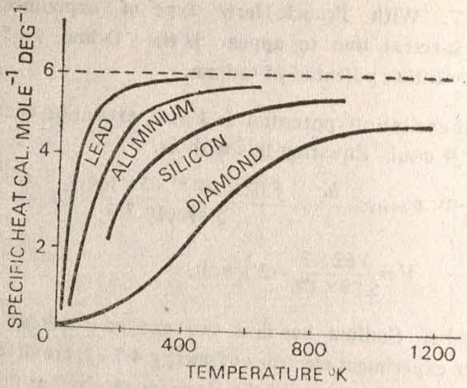


Fig. 19·10.

Variation of atomic specific heats of some simple solids with temperature

Planck's assumption that an oscillator of frequency ν may carry energy only in steps 0, $h\nu$, $2h\nu$, $3h\nu$ etc. was applied by Einstein to the oscillators in the solid state. The average energy for such oscillators is not kT but is given by Eq. (19·6). This replacement leads to $C_v<3R$ if we have $h\nu\gg kT$. Einstein stated that for *C*, *Si* etc., the ν value may be such that at room temperatures $h\nu>kT$. He chose a characteristic $\nu=\nu_E$ (called Einstein frequency to fit the observed sp. ht. verus T curve for the concerned solid. For the solids satisfying Dulong and Petit's law ν_E is such that $h\nu_E$ is less than kT at room temperature ; but as T is decreased kT becomes comparable with and then less than $h\nu_E$. He could fit the theoretical C_v with the experimental curves (Fig. 19·10) by choosing the appropriate ν_E for each solid.

Thus the quantum assumption of Planck was used by Einstein to solve this great riddle which classical physics failed to explain. Of course, many finer details are now added to Einstein's simple theory. There was no independent evidence for the existence of ν_E. Later theories considered a distribution of ν values (for which there is evidence) and calculated sp. ht. from them for different T. But the essence is in Einstein's acceptance of Eq. (19·6) for ε in place of kT.

A similar problem existed in the specific heats of diatomic gases. Again the quantum assumption $\epsilon=nh\nu$ for vibrational energy explained the experimental results satisfactorily, ν in this case being known from spectroscopic data.

19·8. General Comments. It may be noted that we have not made any attempt to explain *why* the energy in electromagnetic radiations should occur in bundles of size $h\nu$ or *why* the energy of an oscillator (whether atoms in a solid lattice or molecules) should occur in steps of $h\nu$ or *why* an atom in free state can have only discrete energy states. That does not amount to any cheating. Any new theory starts with some postulates (*i.e.*, assumptions which are not explained), and Planck and Einstein used the quantum postulate to explain the experimental facts. If these postulates are integrated into a regular mechanics we would stand on a firm ground. For example, we cannot just use Newtonian mechanics to deduce the frequency ν of an oscillator and then impose a statement that energy may ocuur only in multiples of $h\nu$. Again we cannot use Maxwellian electromagnetism to explain electromagnetic oscillations and waves and then impose the statement that energy somehow is bundled up in quanta of magnitude $h\nu$.

We have not dealt with the regular theory—now called quantum mechanics—but have only seen how certain experimental facts unexplained by classical mechanics can be explained if the new mechanics would have scope for the quantization of the energy in electromagnetic wave, in oscillators and in atoms. In fact that is how Physics stood between the years 1900 and 1933 or so. More and more phenomena could be explained (and even predicted) on such *ad hoc* assumptions added to classical mechanics, but no integrated mechanics came up. Quantum mechanics as a regular theory was developed by Heisenberg, Schrodinger and Dirac between 1925 and 1936 in rapid strides after an initial step by de Broglic in 1924. But we shall not go father at this stage.

PROBLEMS

19·1. Given that for the solar surface $T=6000K$ and e_λ is maximum at $\lambda=5500A$, deduce (*i*) λ_{max} corresponding to 300 K, and (*ii*) temperature corresponding to $\lambda_{max}=4000\ A$.

19·2. Planck's theory gave the unknown function $F(\lambda T)$ of Wien's formula. Justify this statement. Did Planck's theory justify the law of Rayleigh and Jeans ?

19·3. A 100 g mass oscillates with frequency 20 *Hz* and amplitude 6 mm. Consider Planck's postulate for it, and comment whether the application is not valid, or the observation of discreteness in energy would be impracticable.

19·4. Light from a 10 milliwatt source at 5 metre distance falls on a surface. On classical wave theory compute the time required for an atom of the surface to gather 5 eV energy needed to expel an electron, assuming that the atom gathers energy falling upto 10 atomic diameters ($\sim 10A$) around it. Discuss how photoelectric effect disagrees with this picture.

19·5. Discuss the difference between Planck's postulate and Einstein's postulate about quantization. Comment on the statement that the emergence of quantum nature of light is *not* a return to Newton's corpuscular concept.

19·6. State the chief *experimental* results of Millikan's work on the photoelectric effect. Deduce Einstein's postulates from these results.

19·7. Light of wavelength λ and intensity I ejects, from a given surface, electrons whose maximum speed is v. How will v be affected if the parameters are separately changed as follows : (*i*) λ is decreased to half value, (*ii*) I is increased 10-fold, (*iii*) the surface is changed ?

19·8. Light of intensity $1{\cdot}2\times10^{-2}$ Wm^{-2} falls on a photocell of surface area $0{\cdot}43\times10^{-3}$ m^2. If $\lambda=4500A$, work function $2{\cdot}00$ eV, and efficiency of photoemission 15%, deduce (*i*) the kinetic energy of the fastest photoelectrons, and (*ii*) the satutation photocurrent.

19·9. The work-function of a given cathode surface is $2{\cdot}30$ eV. Deduce (*i*) the maximum wavelength of light for which photoemission of electrons occurs, (*ii*) the speed of electrons emitted for light of $\lambda=2537$ A.

19·10. Deduce the difference of kinetic energy of the fastest photoelectrons emitted from a given surface for lights of wavelengths 2537 A and 4358 A.. Does this difference depend upon (*i*) the surface chosen, (*ii*) the intensities of the lights ?

19·11. Critically comment on the following statements :

(*a*) If we look into a cavity whose walls are maintained at a constant temperature no details of the interior are visible.

(*b*) Einstein's theory of photoelectricity invalidates Young's experiment on the interference of light.

(*c*) Atoms in gaseous state have a discrete absorption spectrum, but not sc the atoms of a solid.

19·12. Deduce (*i*) the Compton shift of wavelength, and (*ii*) fraction of energy shared by the recoil electron when a photon of $\lambda_0=0{\cdot}674$ $A°$ suffers Compton scattering by angle 120°.

19·13. Light of intensity I falls normally on a disc of radius r. A fraction R is reflected, the rest is absorbed. Deduce an expression for the force due to the radiation. Repeat, taking a sphere (for the disc), other details remaining the same.

19·14, Why is X-ray region the most suitable for Compton shift studies ? Why do we not get just two sharp peaks in fig. 1·95 though the incident radiation is monochromatic ?

19·15. A beam of light of intensity $5{\cdot}00\times10^{+2}$ Wm^{-2} falls at angle 30° from the normal on a plane sheet which has transmission 40%, reflection 25% and absorption 15%. Calculate the radiation pressure on the sheet.

19·16. A spectral series is expressed by

$$\bar{\nu}_n=\frac{109741}{(2-0{\cdot}061)^2}-\frac{109741}{(n-0{\cdot}24)^2} \qquad (n=3,\ 4\ldots)$$

Calculate the λ corresponding to the first line, and also for the series limit. Calculate the ionization potential if the constant term represents the ground state.

19·17. Four spectral lines of atomic hydrogen fall at λ values 6563, 4861, 4341 and 4102 *A*. Fit them in Balmer's formula, and also deduce Rydberg Constant. What will be the λ of the series limit of this series ?

19·18. In a Franck and Hertz experiment an atomic gas shows excitation potentials 6·7 volt and 8·9 volt. If electrons accelerated through 14·0 volt are shot into a gas containing these atoms, deduce (*i*) the possible energies of electrons reaching the collector, (*ii*) the possible wavelengths of the emitted light.

19·19. An atom has the first excitation potential at 3·27 volt and its isonisation potential is 10·41 volt. Deduce (*i*) the binding energy of the outermost electrons in ergs, (*ii*) the longest wavelength in its absorption spectrum.

19·20. If the first three absorption lines of an atom are at 6000, 4000 and 3600 *A*, draw the energy level diagram to scale and show the transitions for these absorptions. Compute also all the *conceivable* lines in emission (not all of which may really occur).

19·21. Show that on Einstein's theory the specific heat of a solid at temperature T will be $x^2\,(e^x-1)^{-2}$ time the Dulong and Petit value, where $x=h\nu_E/kT$ and ν_E is Einstein frequency for that solid. Compute the factor for $x=0{\cdot}5$, 1 and 2.

19.22. "Intensity of light of a given λ falling on a surface is increased 4-fold." What would this statement mean on (*i*) the classical wave theory, and (*ii*) quantum theory ? Under what conditions would an experiment decide between the two views ?

20

The Shell Model of the Atom

Atoms give characteristic spectra as discussed in § 19·4. The energy states they can take are quantized as seen in the explanation of Ritz principle through energy level diagrams (§ 19·5) and by the direct experiments of Franck and Hertz type (§ 19·6). Now, what is the structure of atoms and how may discrete energy levels arise ? This Chapter deals with evidence related to there questions.

Through J.J. Thomson's experiments on cathode rays it was known around the year 1900 A.D. that all atoms contained electrons and an equal amount of positive charge. Also, through molar density of liquids/solids and Avogadro number N, the size of an atom was settled to be a few Angstroms across. J.J. Thomson had given a water-melon model of the atom in which the electrons were like the seeds and the positive charge was spread over the entire atomic volume, in which the seeds were embedded. But direct experiments proved otherwise.

21·1. Experiments on the Scattering of α-particles. In the years 1911 to 1913, Geiger and Marsden (under the guidance of

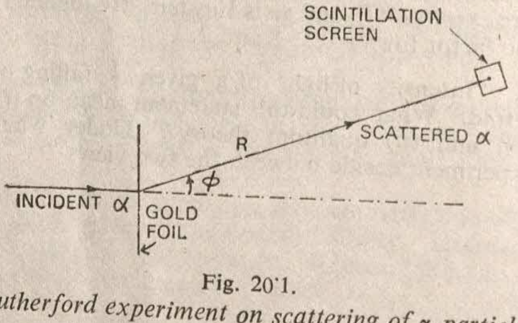


Fig. 20·1.
Rutherford experiment on scattering of α-particles.

Rutherford) did a most remarkable series of experiments to study the structure of atoms. They shot α-particles through a thin foil of gold, and observed the number of them deflected through different angles

ϕ. Fig. 21·1 shows the arrangement schematically. The α-particles came from the radioactive element polonium, their speed being $1{\cdot}6\times10^9$ cm/sec*. The scattered α-particles were received on a screen at various angles ϕ at a distance R. They caused scintillations in the fluorescent material on the screen, and a microscope was used to count the scintillations per minute in a fixed field of view. The whole arrangement was kept in vacuum to avoid collisions of α-particles with air molecules on the way. The gold foil was kept very thin, being $\sim10^{-5}$ cm in thickness, to ensure that whatever deflection was observed was due to single encounters of α-particles with gold atoms.

The most important aspects of the observations were the following :

(*i*) A very large proportion of the α-particles passed through the foil with negligible angular deflection.

(*ii*) A small but finite number of α-particles suffered quite large angle scattering, including 180° deflection.

(*iii*) The detailed angle-wise distribution of scattered particles followed a $\operatorname{cosec}^4(\phi/2)$ law.

Rutherford deduced valuable inferences from these facts about the structure of the atoms of gold. We proceed to outline them.

(*a*) Gold atoms have diameter $d\sim2{\cdot}6\times10^{-8}$ cm†, so that in the 10^{-5} cm thick gold foil used in the experiments there are thus $\sim$400 layers of gold atoms. The fact that a very large proportion of the α-particles went with negligible deflection would mean that the massive part of the gold atoms was limited to a diameter much less than the atomic diameter d.

(*b*) Large angle scattering of some α-particles means that they face a potential barrier of height exceeding the kinetic energy with which they come. Rutherford considered that if the positive charge Ze was concentrated in a sphere of radius r, with Z electrons outside somewhere, the maximum potential energy of the alpha particle and positive charge Ze system would be (assuming inverse square law force)—

$$U=\frac{2e.Ze}{r}=\frac{2Ze^2}{r} \qquad \ldots(20{\cdot}1)$$

Now, if an α-particle is to rebound (180° scattering), we have

$$\frac{2Ze^2}{r}>\tfrac{1}{2}m_\alpha V^2 \qquad \ldots(20{\cdot}2)$$

---

* α-particle has mass $\sim$4 hydrogen atoms and charge +2 electronic charge. In radioactivity they come off spontaneously, and their speed is characteristic for each radioactive material.

† $d\approx(\text{volume})^{1/3}\approx(\text{mass/density})^{13}\approx[1{\cdot}97\times1{\cdot}67\times10^{-24}/19{\cdot}3]^{1/3}$.

For the α particles from polonium $V=1{\cdot}6\times10^{9}$ cm/sec., and from X-ray scattering experiments Z for gold is given as 79. Substitution in Eq. (20.2) leads to

$$r<4\times10^{-12}\ \text{cm}.$$

Thus the positive charge of the gold atom must be concentrated in a radius roughly 10^{-4} times the atomic diameter.

Notice that we are not asking how many particles suffer large angle scattering; just the fact that such scattering occurs gives us an upper limit for r. Later on α-particles of about 10-fold energy have also been found to suffer back-scattering, which sets the upper limit of r to a correspondingly lower value. We may thus write broadly

$$r\sim10^{-12}\ \text{cm}.$$

In Eq. (20·1) we have assumed that the inverse square law of electrostatic forces remains valid at $r\sim10^{-12}$ cm. If a different law applied at such small distances, the r value deduced would not be valid.

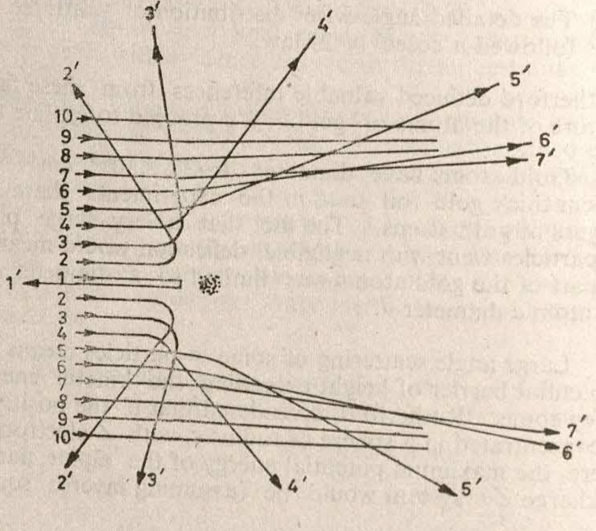


Fig. 20·2.

Schematic representation of deflection of α-particles by a nucleus.

(c) Rutherford next used the inverse square law of force to deduce the angular distribution of the scattered α-particles. Fig. 20·2 gives a schematic representation of a stream of α-particles passing by a nucleus. Those particles which would pass at a closer lateral distance from the nucleus are deflected more ; those (like 1 in Fig. 20·2) proceeding head-on towards the nucleous suffer 180° deflection; those (like 10 in Fig. 20·2) passing far from the nucleus suffer little deflection.

If N_0 is the number of incident particles per cm^2 per second, and N_ϕ is the number received on a screen of given size per second in a given direction ϕ, the inverse square law gives the result

$$N_\phi = N_0 K (\operatorname{cosec} \phi/2)^4 \qquad \ldots(20{\cdot}3)$$

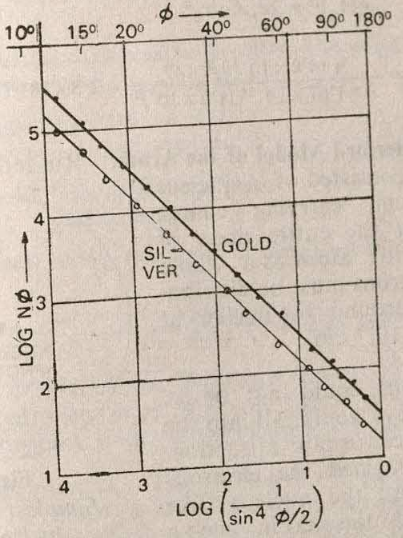


Fig. 20·3.

Quantitative verification of the applicability of Coulomb's Law in α-scattering

where K is a constant proportional to Z^2 and V^{-4} and a factor depending on the geometry of the observing system.

The experiments of Geiger and Marsden showed very good agreement with eq. (20·3), as shown in Fig. 20·3. (Later experiments have shown that the constant K is proportional to Z^2 and V^{-4}, both of which agree with the theoretical predictions from Coulomb's law).

Thus this series of careful experiments firmly established that the mass and positive charge of the atoms is concentrated in a nucleus of size $\sim 10^{-12}$ cm. and that Coulomb's law is applicable for distances as small as this.

In fact certain departures from the results expected on simple Coulomb interaction are observed. But we shall not go into these finer points. One may imagine that projectiles of greater kinetic energy will go nearer the nucleus and then the departures may become more and more serious. The matter is not that simple. More serious departures may occur at particular energy values than at higher energy values, meaning that some kind of 'resonances' occur.

Example 20·1. With α-particles from polonium ($v=1{\cdot}6\times10^9$ cm/sec) calculate the nearest approach r_0 with silver nuclei ($Z=47$).

Solution. Nearest approach is for head-on collision, and all kinetic energy then becomes potential energy.

$$\tfrac{1}{2}m_{\alpha}v^2=2e.Ze/r_0{}^*$$

$$r_0=\frac{4\,(4{\cdot}8\times10^{-10})^2\times47}{4\times1{\cdot}67\times10^{-24}\,(1{\cdot}6\times10^9)^2}=2{\cdot}5\times10^{-13}\text{ cm.}$$ ⊙

20·2. Rutherford Model of the Atom. Rutherford established that the atom consisted of a nucleus of size $\sim10^{-12}$ cm, carrying charge $+Ze$, and almost the entire mass of the atom. Since the atom as a whole is neutral, Z electrons must be distributed *in some way* around the nucleus at distances upto $\sim10^{-8}$ cm.

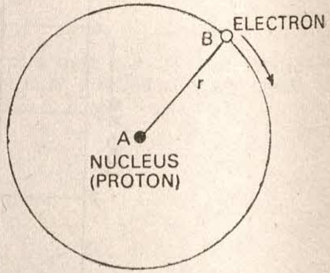


Fig. 20·4.
Simplest model of the hydrogen atom

The electrons could not be at rest, otherwise they would fall into the nucleus due to electrostatic attraction. So Rutherford imagined the electrons as revolving around the nucleus (like the planets round the sun) in such a way that the electrostatic attraction provided the centripetal force. This is called the *Rutherford model* of the atom (Fis. 20·4).

From elementary electrodynamics we know that an electrical charge going round a circular path with frequency f should emit electromagnetic wave having frequency $\nu=f$. In other words, by arttributing different orbits to different electrons in an atom we could relate them with the different frequencies observed in the emission pectrum. Further, if the electron motions are not simple harmonic, then by Fourier theorem we could explain the occurrence of higher *harmonics* 2ν, 3ν, etc. and also of *cominbations* $\nu_1+2\nu_2$, $2\nu_1+3\nu_3$, etc. However, observed frequencies in the spectra of atoms do not show such relationship at all, as seen in Chapter 19.

There is an even more serious difficulty. As the system loses energy by radiation, the orbit of the moving charge must go on contracting and the frequency must go on increasing continuously. Thus the classical theory would account neither for sharp lines in the spectra nor for the stability of atoms.

---

*'Nearest approach' distance r_0 is greater than the nuclear radius r. Notice that we use equality sign here, unlike in Eq. (20.2).

It was in these circumstances that Bohr (1913) put forth his bold postulates to quantize the energy states of the atom, just as Planck had done 13 years earlier to quantize the energy in oscillators in the walls of a blackbody chamber.

Example 20·2. At what frequency should an electron revolve around a proton if orbit radius is 0·53 *A* ? Estimate the wavelength of radiations emitted on the classical theory.

Solution. Equating centripetal force with the Coulomb force,

$$mv^2/r=e^2/r^2 \Rightarrow v=e/(mr)^{1/2} \qquad \text{...(20·4)}$$

Frequency is given by $v/2\pi r$. Hence,

$$f=v/2\pi r=[e/2\pi]m^{-1/2}\, r^{-3/2} \qquad \text{...(20·5)}$$

With $r=0{\cdot}53\times10^{-8}$ cm, substitution gives

$$f=6{\cdot}6\times10^{15}\ s^{-1}$$

On the classical theory this is also the frequency of radiation emitted. Hence

$$\lambda\equiv\frac{c}{f}=\frac{3{\cdot}0\times10^{10}\ \text{cm}\ s^{-1}}{6{\cdot}6\times10^{15}\ s^{-1}}\approx 4{\cdot}5\times10^{-6}\ \text{cm}\ (450\ A^\circ)^*$$

20·3. Bohr Model of the Hydrogen Atom. To explain both stability of the hydrogen atom and the series of spectral lines in emission and absorption, Bohr put forth the following postulates :

1. *Stationary Orbits.* The motion of the electron in the field of the nucleus is *restricted to one of a certain set of stationary orbits.* While moving in any one such orbit, it does *not* radiate out energy.

2. *Quantum Condition.* The stationary orbits are given by the rule that *angular momentum* p_ϕ of the electron about the nucleus is :

$$p_\phi\equiv mvr=\frac{nh}{2\pi},\ (n=1,2\ldots) \qquad \text{...(20·6)}$$

in which h is Planck's constant and n is an integer, called the quantum number of the given state.

3. *Transitions.* If the electron jumps from a higher to lower energy orbit, the difference of energy is given out as a packet of radiation. If the jump is from a lower to a higher energy orbit, the required packet of energy must be absorbed from some external agency.

4. *Frequency Condition* : If E_i and E_f are the energies of the two states involved in a transition, then the radiation given out or absorbed in the transition has a definite frequency ν given by the condition

$$\nu=\frac{|E_i-E_f|}{h}. \qquad \text{...(20.7)}$$

where h is again Planck's constant.

***Actually $\lambda=c/\nu$, and assumption $\nu=f$ is not valid in Bohr theory.**

With these assumptions, Bohr's model explains the spectrum of atomic hydrogen in the following steps :

(*a*) *Energy expression from classical theory* : For a circular n^{th} orbit of radius r_n traversed with speed v_n, we equate centripetal acceleration with electrostatic attraction. For a nuclear charge Ze, we have

$$\frac{mv_n^2}{r_n}=\frac{Ze^2}{r_n^2} \Rightarrow r_n=\frac{Ze^2}{mv_n^2} \qquad ...(20{\cdot}8)$$

This gives kinetic eaergy of the electron : $E_{kin}=\frac{1}{2}mv_n^2=Ze^2/2r_n$. Adding the potential energy $-Ze^2/r_n$ with respect to the state of ionized atom as zero level, the total energy is given by

$$E=E_{kin}+E_{pot}=-\tfrac{1}{2}Ze^2/r_n \qquad ...(20{\cdot}9)$$

(*b*) *Quantization of the orbits* : Bohr's quantum condition for the n^{th} orbit gives $v_n=nh/(2\pi mr_n)$. Substitution in Eq. (20.8) and solving for r_n gives

$$r_n=\frac{n^2h^2}{4\pi^2e^2mZ}=a_o\frac{n^2}{Z}\ ; \quad \text{where} \quad a_0=\frac{h^2}{4\pi^2e^2m} \qquad ...(20{\cdot}10)$$

The quantity a_0 (called *Bohr radius*) is the radius for $n=1$ and $Z=1$. From known constants we get $a_o=0.5268$ A°.

(*c*) *Quantization of energy* : Eqs. (20.9) and (20.10) give

$$E_n=-\frac{2\pi^2e^4mZ^2}{n^2h^2}=-\frac{RhcZ^2}{n^2}=-\frac{13{\cdot}61Z^2}{n^2} \text{ electron-volt} \qquad ...(20{\cdot}11)$$

$$\text{where} \quad R\equiv\frac{2\pi^2e^4m}{h^3c}, \text{ (Rydberg Constant)}=1{\cdot}097\times10^5 \text{ cm}^{-1} \qquad ...(20{\cdot}12)$$

(*d*) *Emission spectra* : If suffixes i and f refer to the initial (higher) and final (lower) states, respectively, then the frequency ν of light emitted in this transition is given by

$$\nu=\frac{E_i-E_f}{h}=-\frac{Rc}{n_i^2}-\left(-\frac{Rc}{n_f^2}\right)=Rc\left(\frac{1}{n_f^2}-\frac{1}{n_i^2}\right)$$

The reciprocal wavelength (wave number) of the radiation is given by

$$\bar{\nu}=\frac{\nu}{c}=R\left(\frac{1}{n_f^2}-\frac{1}{n_i^2}\right) \qquad ...(20{\cdot}13)$$

If we put $n_f=1$ and let n_i be a running number >1...we get a series

$$\bar{\nu}=R\left(\frac{1}{1^2}-\frac{1}{n^2}\right), \qquad (n=2, 3, 4...) \qquad ...(20{\cdot}14)$$

This has the same form as eq. (19.17 *a*) for the Lyman series. We may similarly obtain expressions for the other observed series in the spectrum. Note also that R (calculated) $= R$(experimental).

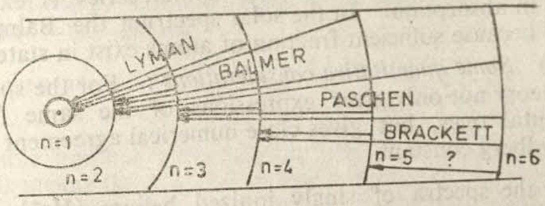


Fig. 20'5.

Representation of electron orbits in hydrogen atom, with the transitions for emission

A schematic representation of the various Bohr orbits is given in Fg. 20'5 and the corresponding energy levels are shown in Fig. 20'6.

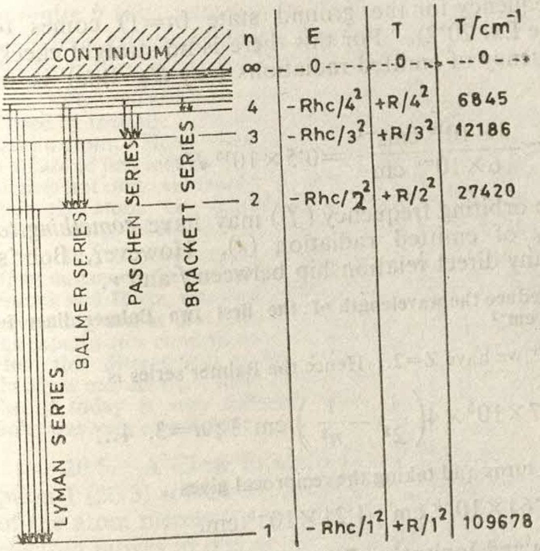


Fig. 20'6.

Energy level diagram for the H atom

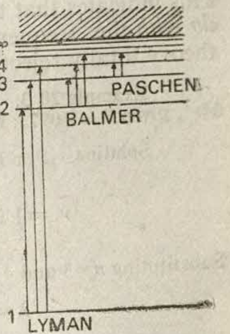


Fig. 20'7.

Transitions for absorption : Only the Lyman series (L) is observed at ordinary temperatures

The transitions for different spectral series in emission are shown in both these figures. The quantity 'term value' is given by $-E_n/hc$, and differences of term values directly give the $\bar{\nu}$ value of the related emission or absorption line.

(*e*) *Absorption spectra.* Absorption corresponds to jumps from a lower to a higher energy orbit (Fig. 20'7). However, at ordinary temperatures almost all the atoms in atomic hydrogen are in the ground state $n=1$*. Hence only the Lyman series is experimentally observed in absorption. In the solar spectrum the Balmer series is observed because sufficient fraction of atoms exist in state $n=2$ also.†

(*f*) *Some quantitative considerations* : For the spectral lines, Bohr's theory not only gives expressions of the same *form* as the experimental ones, but gives close numerical agreement in the value of the Rydberg constant.

For the spectra of singly ionized helium (He^+) and doubly ionized lithium (Li^{++}) the same equations with $Z=2$ and 3 respectively should represent the spectral series. This is found to be the fact.

$a_0=0{\cdot}53\ A$ gives an order of magnitude $1A^\circ$ for the atomic diameter of H atom, which is not in conflict with the order estimated from other considerations like viscosity, conductivity, etc.

The orbiting frequency for the ground state ($n=1$) comes to $f=6{\cdot}6\times10^{15}\ Hz$ (See Ex. 20·2). For this there is no straight check. Incidentally, the frequency of emitted radiations (say H_α line) is of the same order :

$$\nu=\frac{c}{\lambda}\approx\frac{3\times10^{10}\ \text{cms}^{-1}}{6\times10^{-5}\ \text{cm}}=0{\cdot}5\times10^{15}\ Hz$$

This indicates that the orbiting frequency (f) may have *something to do with* the frequency of emitted radiation (ν). However, Bohr's theory totally rejects any direct relationship between f and ν.

Example 20·3. Deduce the wavelength of the first two Balmer lines in He+, given R=1·097 × 10⁵ cm⁻¹

Solution. For He^+ we have Z=2. Hence the Balmer series is

$$\bar{\nu}=1{\cdot}097\times10^5\times4\left(\frac{1}{2^2}-\frac{1}{n^2}\right)\text{cm}^{-1}\ ;\ n=3,\ 4...$$

Substituting $n=3$ and 4 in turns and taking the reciprocal gives

$$\lambda=1{\cdot}64\times10^{-5}\ \text{cm}\ ;\ 1{\cdot}21\times10^{-5}\ \text{cm}.$$

20·4. Excitation and Ionisation Potentials of Hydrogen.**

The energy levels are often expressed in the unit *electron volt*, abbreviated *ev*. Bohr's theory [Eq. 20·11] leads to

$$E_n=-\frac{13{\cdot}61}{n^2}\,ev.$$

---

[*] If E_{12} is energy difference between states 1 and 2, the ratio of atoms in the two states at temperature T is given by Boltzmann law in thermodynamics as $N_2/N_1=\exp\left[-E_{12}/kT\right]$. For atomic hydrogen levels with $n=2$ and $n=1$, this ratio comes to exp (—400) at room temperature, and exp (—20) at the temperature of the Sun.

[†] Lyman region is suppressed by absortion by earth's atmosphere.

[**] See also § 19·6, Franck and Hertz experiments.

The *excitation energy* for any state n is measured relative to the ground state ($n=1$). Hence

$$\text{excitation energy}=-13{\cdot}61\left(\frac{1}{n^2}-\frac{1}{1^2}\right)=13{\cdot}61\left(\frac{1}{1^2}-\frac{1}{n^2}\right)ev \qquad \ldots(20{\cdot}15)$$

The same number expressed in volts is called the *excitation potential*. It means the *potential* through which an electron should fall so as to acquire that much *energy*. By putting $n=2, 3\ldots\ldots$ we get the successive excitation potentials as 10·21, 12·10......volt.

If we put $n=\infty$ in equation (20·15) we get what is called the *ionisation potential of atomic hydrogen*. Its value is 13·61 volt. It is *the potential through which an electron must be speeded up so that on colliding with an atom of hydrogen it is just able to ionise it* (*i.e.*, separate its electron and proton with zero kinetic energy). The ionisation *energy* for hydrogen atom is 13·61 electron-volt.

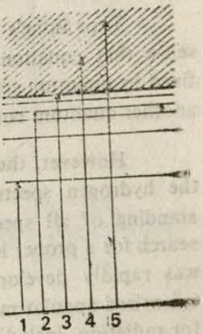


Fig. 20·8.

Fig. 20·8 shows some transitions which may be caused by inelastic collisions between an external agency and an atom. Here, 1 and 2 are cases of 'excitation', 3 is a case of just ionisation, and 4 and 5 are cases where there is not only separation of the outermost electron from the atom, but there is extra energy which goes as kinetic energy of free motion of the separated particles.

In § 19·6 we described an experiment which in effect measured the excitation potentials. That work by Franck and Hertz, was published just one year after Bohr gave his theory. (It is interesting that at that time they were not aware of Bohr's theory). However, the experiments confirm only the concept of discrete energy states of an atom. How this discreteness arises is not answered by these experiments. Bohr's theory is one answer ; there may be other answers too ! In fact the quantum theory today is very different from how Bohr put it. But the initial step by Bohr was very significant.

20·5. A Close Examination of Bohr's Theory. Bohr's assumption 1 (20·3) solved the problem of stability of the Rutherford model of the atom merely by *postulating* that there is no emission while the electron moves in one of the allowed orbits. It was a drastic postulate, which threw away the only mechanism of radiation of electromagnetic waves known at the time. It violates the classical idea : an electron in circular orbit is an accelerated charged particle and it must radiate if classical theory is to remain valid.

Assumption 2 appears arbitrary. Why the angular momentum should be quantised and not some other property like v or r or E? And why the quantum step should be in units of $h/2\pi$? However, Planck had earlier discussed quantization of the harmonic oscillator (Chapter 19). Its corresponding form for the present case would be

$$\int_0^{2\pi} p_\phi \, d\phi=nh \;. \quad \text{(n integar)} \qquad \ldots(20{\cdot}16)$$

which gives Eq. (20·6). Thus Bohr's assumption is in harmony with Planck's approach for a harmonic oscillator.

Assumption 3 is silent about the *mechanism* by which radiation occurs; it merely tells that if there is a downward transition, the energy difference comes out as radiation. Now, from a given higher energy state the atom can pass on to several lower energy states. Bohr's assumption does not say to which of these states the atom will pass spontaneously, at what time, and in what preference or proportion.

Assumption 4 is again a bold postulate, but since Einstein had used it successfully in another context (photoelectric effect, Chapter 19), it was not altogether out of the way. Eq. 20·7 is called the Einstein frequency condition.

Thus Bohr's theory is a hybrid of classical and quantum ideas in the sense that equation (20·10) is obtained from classical ideas, while the orbits are fixed by quantum condition (20·6) and the frequency of radiation is fixed by another quantum condition (20·7), rejecting $\nu=f$.

However, the theory was astonishingly successful in explaining not only the hydrogen spectrum quantitatively, but in paving the way for an understanding of all spectra and atomic structures. Therefore, there was frantic search for a proper logical basis for the assumptions made. Quantum Mechanics was rapidly developed. That theory follows logical lines, leads naturally to quantized angular momenta and quantized energy states, and gives a mechanism for radiation, including quantitative values of transition probabilities between different energy states. That theory (the quantum mechanical theory) does not talk of definite electron orbits, radii, velocity, etc., but only of definite energy states. It is logically far more sound and statisfying. But its ideas are rather abstract. Therefore, Bohr's simple picture, though logically less sound, is still of great interest for the beginner.

20·6. The Shell Model of the Atom. Bohr's theory was extended and refined by several workers to explain ; (i) finer splittings within each spectral line of the hydrogen spectrum, and (ii) the structure, spectra and physical and chemical properties of other atoms. Ultimately a fundamentally different approach (*viz.*, quantum mechanics) was developed, which broadly gave the same results as Bohr's theory, but went much farther in giving many other results. We shall only give the outlines of some significant results from quantum mechanics.*

---

* See § 23·9 for details.

(a) Quantum numbers n, l and m

The quantum mechanics, applied to electron+nucleus system, gives three quantum numbers :

- n, *the total quantum number*, which governs the major part of the energy through terms like $E_n = -R/n^2$
- l, *the azimuthal quantum number*, which governs the orbital angular momentum $L_l = lh/2\pi$,† but also slightly affects the energy. Thus each E_n level becomes a multiplet level described by E_{nl}. For a given n the allowed values of l are 0, 1, 2,...$(n-1)$.
- m_l *the magnetic‡ quantum number for l*, which governs the component of orbital angular momentum along a specified axis : $L_z\,(l) = m_l h/2\pi$. For a given l the allowed values of m_l are from $+l$ to $-l$ in steps of 1.

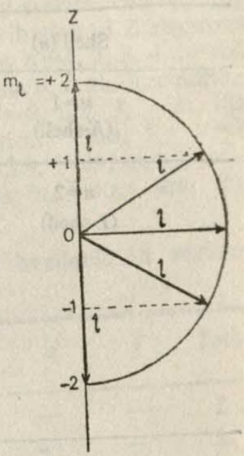


Fig. 20·9.
Space quantization for l=2.

In Fig. 20·9 we show the case of $l=2$. Imagine an orbit with its plane ⊥′ to the plane of paper. Then the vector $lh/2\pi$ lies in the plane of paper. However, if z is a defined axis, then l can take just five possible orientations with respect to z-axis, such that the component of $lh/2\pi$ along z can take five possible values $(+2, -1, 0, -1, -2)h/2\pi$.

(b) Electron Spin

In addition to the above, the electron itself is attributed an intrinsic spin angular momentum $sh/2\pi$, with quantum number s having a fixed value $\frac{1}{2}$. The corresponding space quantization of electron spin gives us

m_s, *the magnetic quantum number for s*, which governs the component of spin angular momentum along a specified axis : $L_z(s) = m_s h/2\pi$. Since $s = \frac{1}{2}$, there are just two possible values of m_s : + 1/2 and −1/2.

---

† Note that in Bohr's theory the *angular momentum* is also determined by quantum number n)

‡ A magnetic moment is associated with l as well as s. In a magnetic field the components of these along the Z axis determine the magnetic contributions to energy. Hence quantum numbers m_l and m_s bear the adjective 'magnetic'.

Table 20·1. Available States for Different n Quantum Numbers in an Atam

Shell (n)	Sub-shell (l)	m_l	m_s
$n=1$	$l=0$	0	1/2
(K-shell)	(1s)	0	−1/2
$n=2$	$l=0$	0	1/2
(L-shell)	(2s)	0	−1/2
	$l=1$	1	1/2
	(2p)	1	−1/2
		0	1/2
		0	−1/2
		−1	1/2
		−1	−1/2
$n=3$	$l=0$	0	1/2
(M-shell)	(3s)	0	−1/2
	$l=1$	1	1/2
	(3p)	1	−1/2
		0	1/2
		0	−1/2
		−1	1/2
		−1	−1/2
	$l=2$	2	1/2
	(3d)	2	−1/2
		1	1/2
		1	−1/2
		0	1/2
		0	−1/2
		−1	1/2
		−1	−1/2
		−2	1/2
		−2	−1/2

(c) Coupling of orbit and spin momenta

In any multi-electron atom the orbital as well as spin angular momenta of different electrons are *coupled* together following some quantum rules again. But we will not go into these coupling schemes. For a single electron (as in H atom) the result of l and s coupling gives a j quantum number, which can take only two values $l+1/2$ and $l-1/2$ (since $s=1/2$). However, in a strong magnetic field these couplings break, so that j loses meaning. Only n, l, m_l and m_s then define uniquely the 'state' of each electron.

(d) Pauli's exclusion principle

It says that no two electrons in a given system can occupy identical states*. This principle decides the distribution of Z electrons of an atom in successive energy shells defined by $n=1, 2, 3, 4...$ and called *K*, *L*, *M*, *N*...shells. Table 20·1 shows the n, l, m_l, n_s combinations possible (upto $n=3$). The sub-shells $l=0, 1, 2, 3,$ bear the symbols *s*, *p*, *d*, *f*... from spectroscopic practice. Thus, for $n=3$ (*M*-shell), you can have 3 sub-shells with $l=0$ (3*s*), 1 (3*p*) and 2 (3*d*). Then for a given sub-shell you can have various m_l values and for each of them m_s could be $+1/2$ or $-1/2$.

The procedure gives the number of states available in various shells and subshells as follows :

Shell	*Sub-shell* : *s*	*p*	*d*	*f*	*Total*
$n=1$ (*K*) :	2	—	—	—	2
2 (*L*) :	2	6	—	—	8
3 (*M*) :	2	6	10	—	18
4 (*N*) :	2	6	10	14	32

(f) Electronic Configuration of an Atom.

The Z electrons of an atom start filling from the lowest energy states, going upwards. The order of energies is decided by two rules : (i) A lower n means lower energy, and (ii) within a given n, a lower l means lower energy. However, if n goes up by one step and l goes down by two steps, we reach a lower energy level†.

Keeping these in mind the levels fill successively in the order 1*s*, 2*s*, 2*p*, 3*s*, 3*p*, 4*s*, 3*d*, 4*p*, 5*s*, 4*d*, 5*p*......As an example the distribution of 11 electrons of normal sodium atom may be described in the form given below :

$$(1s^2\ 2s^2\ 2p^6)\ 3s^1$$

where the principal quantum number n is written in the main line, azimuthal quantum number l is expressed by *s*, *p*, etc. and the number of electrons in each n, l state is written as superscript. The bracket only shows the complete shells (or sub-shells), so that one electron in 3*s* state ($n=3$, $l=0$) is in the outermost uncompleted shell. The above description is called *electronic configuration* of the atom. As another example, the electronic configuration of calcium atom in the ground state is

$$(1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6)\ 4s^2$$

---

* This is a general rule following from quantum mechanics for all spin-half particles, which are also called Fermions or Fermi-Dirac particles. Integral spin particles are, in contrast, called Bosons (named after SN Bose) and any number of them can occupy a given state.

† As an example, from 3*d* to 4*s* we find n increasing by *one* step and l decreasing by *two* steps; so 4*s* state is lower than 3*d*.

Note that the $3d$ sub-shell has not started filling; $4s$ fills first, which has l value 2 steps lower than that for $3d$.

It is customary to consider the completed shell as *core*, or the electrons there as *core-electrons.* The outermost electrons are called *valence electrons* in chemistry and *optical electrons* in spectroscopy.

It may be noted from Table 20·1 that each completed subshell has $\Sigma m_l=0$ and $\Sigma m_s=0$. Hence Σm_l and Σm_s values for all Z electrons are equal to the Σm_l and Σm_s values of the valence electrons alone. Thus the orbital and spin contributions to the angular momentum (and hence the magnetic moment), the chemical activity and magnetic properties, all are determined solely by the electrons of uncompleted shells. They are the electrons which determine chemical combinations and optical spectra.

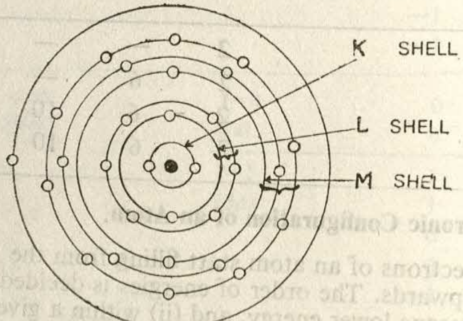


Fig. 20·10.

Schematic representation of the electronic shell structure of copper ($Z=29$), $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s$

Fig. 20·10 gives a schematic representation of the electronic configuration of the normal *Cu* atom. However, this kind of diagram is an over simplification. The relative diameters of shells or the sub-shells must not be taken seriously.

20·7. The Periodic Table of Elements. Pauli's exclusion principle makes the Z electrons in an atom arrange into *K*, *L*, *M*, *N*... shells ; also for each completed shell the angular momentum and magnetic moment become zero (except for nuclear contributions). Hence with increasing Z the properties of elements nearly repeat as each shell is completed. This periodic repetition was observed in chemistry early, whence the periodic table of elements was given.

Consider the electronic configurations of the atoms given below :

H	$1s$	$Z=1$
Li	$(1s^2)\ 2s$	3
Na	$(1s^2\ 2s^2\ 2p^6)\ 3s$	11
K	$(1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6)\ 4s$	19

In each case there is one electron outside the '*core*'. The outermost electron therefore effectively moves under a central charge nearly $+1e$ in all these cases. The atoms with $Z=1, 3, 11, 19$ must have similar properties. They all actually have chemical valency $+1$. They also have similar optical spectra and similar magnetic properties.

Consider next the atoms with the electronic configurations :

He	$(1s^2)$	$Z=2$
Ne	$(1s^2\ 2s^2\ 2p^6)$	10
A	$(1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6)$	18
Kr	$(1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^6)$	36

In each case here the shells (or sub-shells) are complete, with no electrons outside the 'core'. For completed sub-shell $\Sigma m_l=0$, $\Sigma m_s=0$. Thus the 'core' always has net angular momentum (and hence magnetic moment) zero. Hence all these atoms are chemically inert (inactive) and magnetically neutral.

Further, consider the halogens whose configurations are—

F	$(1s^2\ 2s^2)\ 2p^5$	$Z=9$
Cl	$(1s^2\ 2s^2\ 2p^6\ 3s^2)\ 3p^5$	17
Br	$(1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2)\ 4p^5$	35
I	$(1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^2\ 4p^6\ 4d^{10}\ 5s^2)\ 5p^5$	53

Here in each case just one electron more is required to complete the *p* sub-shell. These elements therefore all show chemical valency -1. They have similar optical spectra also.

We can go on discussing the entire periodic table of elements on the basis of electronic configurations. The point is that all chemical, optical, magnetic and other properties of elements are now explained in terms of the electronic configuration of the atoms.

20·8. The Spectra of Alkali Atoms. An alkali atom is characterised by a 'core' of $Z-1$ electrons, and one *optical electron*. Thus the core as a whole has (in alkali atoms) a net charge of $+1e$, and one optical electron is free to move under the field of this charge. Hence, the spectra of alkali atoms are somewhat similar to the spectrum of hydrogen atom. The energy levels are given by terms like those for the *H* atom :

$$E_n=-\frac{R}{(n-\alpha)^2}, \qquad \text{...(20·17)}$$

where n is an integer starting with 2 for *Li*, with 3 for *Na*, and so on, and α is called 'quantum defect' which is positive and depends upon the l quantum number. The reason for deviating from Bohr-like terms $-R/n^2$ and replacing n by $n-\alpha$ may be understood from Fig. 20·11, which is an old depiction for *Na* atom. Apart from the 'core' (shown shaded), three of the lowest states of the optical electron are shown ; $3s$, $3p$, $3d$. All s orbits are very elliptical, and therefore penetrate into the core in part of the paths. These are therefore called *penetrating orbits*. Then are the *p* orbits, which are less elliptical and therefore less penetrating. Then *d* orbits have even less or no penetration.

For the non-penetrating orbit, the nuclear charge *Ze* is shielded by charge $-(Z-1)\,e$ of the core, so that the effective central charge is 1 *e*., But the penetrating electrons, in part of their paths, move under a larger central charge. Hence the penetrating electrons have more binding, *i.e.*, their energy levels are lower (more negative). To account for this penetration, α is introduced in Eq. (20·17). It is largest for *s* states, and is smaller and smaller for *p*, *d*, *f*...states. In sodium the quantum defect is 1·37 for *s*-states, 0·88 for *p*-states, and zero for *d*-states. In rubidium the values are much larger, being 3·19, 2·17 and 1·23, respectively.

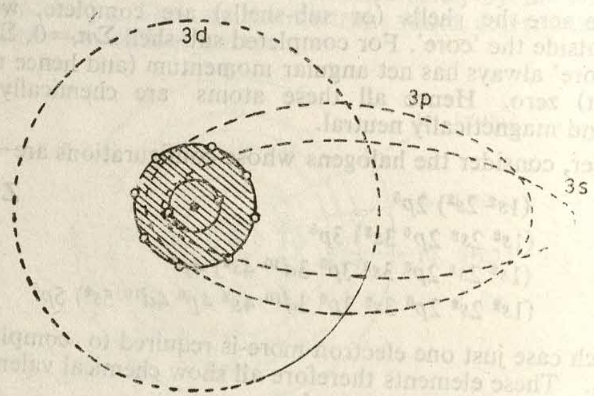


Fig. 20·11.
The core and three of the many possible states of optical electron in sodium (schematic)

Apart from this difference, the spectra of alkali atoms are similar to those for hydrogen atom.

Example 20·4. The *ns* and *np* states of sodium atom have quantum defects 1·37 and 0·88 respectively. Calculate the energy values for the lowest three levels of each set, draw an energy level diagram, and deduce the wavelength of the spectral line for the transition 3p—3s.* (R=109670 cm⁻¹)

Solution. In *Na* the lowest *n* is 3. Using the α values for *s* and *p* states. Eq. (20·17) gives—

$$E_s=-\frac{109670}{(1{\cdot}63)^2},\ -\frac{109670}{(2{\cdot}63)^2},\ -\frac{109670}{(3{\cdot}63)^2}\ \text{cm}^{-1}$$

$$=-41290,\ -15850,\ -8326\ \text{cm}^{-1}$$

---

* This transition gives the familiar *D*-line of sodium, $\lambda=5893\ A°$.

$$E_p = -\frac{109670}{(2{\cdot}12)^2},\ -\frac{109670}{(3{\cdot}12)^2},\ -\frac{109670}{(4{\cdot}12)^2}\ \text{cm}^{-1}$$

$$= -24410,\ -11270,\ -6463\ \text{cm}^{-1}$$

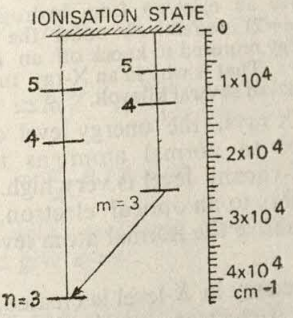


Fig. 20.12.

Energy level diagram for Example 20·4. The s-states are on the left and p states on the right.

The energy level diagram is given in Fig. 20·12. The transition $m=3$ to $n=3$ is also shown. The wavelength is given by

$$\lambda = \frac{1}{\nu} = \frac{1}{41290-24410}\ \text{cm.}$$

$$= \frac{1}{16880} \approx 5{\cdot}9\times10^{-5}\ \text{cm.}$$

20·9. The X-ray Spectra. The spectrum of an X-ray source has two parts : *continuous spectrum*, which has a sharp low λ limit, and *characteristic spectrum*, which has sharp peaks at λ values characteristic of the target (anode) in the X-ray tube. The absorption spectrum of X-rays has also two parts : general absorption at all λ, and characteristic absorption, which gives not absorption lines but *edges*.

The shell structure of atoms is able to explain all these features.

Characteristic X-Rays

In an X-ray tube, accelerated electrons strike a target (say tungsten) and produce X-rays. In an atom with high *Z*, the *K*-shell, *L*-shell, *M*-shell are full, and there may be a vacancy in *N*-, *O*... shell, and (of course) outside the atom. The striking electrons may excite a *K*-shell electron of the target. But since *L*- and *M*-shells are full, Pauli's principle requires that the *K*-shell electron can only go outside the atom (*N*-, *O*-, etc are close to 'outside'). So we get a *K-vacant* state. Similarly, the striking electrons may create a *L-vacant* state or a *M-vacant* state.

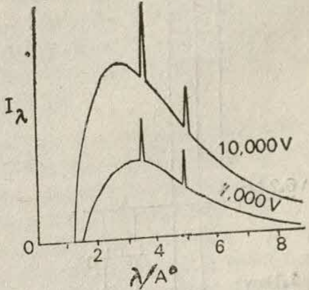


Fig. 20.13.

X-ray emission spectrum from a given source for two voltages.

The process needs high energy. For example in a copper atom ($Z=29$) each K-shell ($n=1$) electron is effectively under the field of charge $\sim 28e$. Hence from eq. (22·10) the radius of K-shell is $\sim 1/28$ of Bohr radius a_0. Also, from eq. (20·11) the binding energy is $\sim (28)^2$ times, that is, the energy required to remove a K-electron *out of this atom* is $\sim [28]^2 \times 13{\cdot}6\ ev$, where $13{\cdot}6\ ev$ is the energy required to remove an electron from hydrogen atom. Similarly each core electron in L-shell ($n=2$) is effectively under the field of charge $\sim (29-9)e = 20e$ and hence the energy required to knock off an L-electron is, by eq. (20·1), about $[(20)^2/2^2)] \times 13{\cdot}6\ ev$. That is why in an X-ray tube free electrons are accelerated through a potential of several kilovolt.

In the case of X-rays, the energy level diagram (Fig. 20·14) is drawn with reference to normal atom as the zero state. Since K-level is deep, the K-vacant level is very high. The L-vacant level is less high. If N refers to an optical electron, then N-vacant level will be just a few ev above the normal atom level, equal to ionization energy.

Now, once a vacancy in K-level is created, it may be filled by a jump of L-, or M-, or N-electron into it. Correspondingly the atom changes from K-vacant to either L-vacant, or M-vacant, or N-vacant state. In Fig. (20·14) the three arrows on the left show these transitions. The emission lines are called the K-series and marked $K\alpha$, $K\beta$, $K\gamma$. The two arrows on the right show transitions in which a vacancy in L-shell is filled by the jump of an M or N-shell electron into it. They are members of L-series.

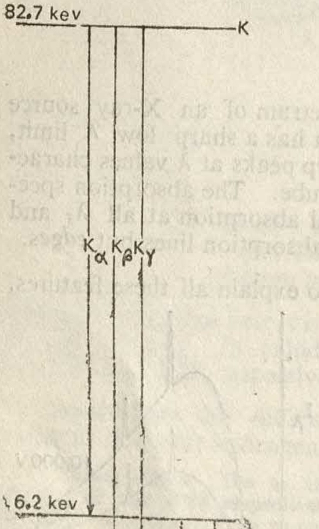


Fig. 20.14.
Energy level diagram for X-ray excitation in gold.

In gold we would expect $K\alpha$ and $K\beta$ lines at $E_K - E_L = 66{\cdot}5$ kev and $E_K - E_M = 78{\cdot}6$ kev ($\lambda \approx 0{\cdot}19$ and $0{\cdot}16A$). Similarly $L\alpha$ line would fall at $E_L - E_M = 12{\cdot}1$ kev ($\lambda \approx 1A°$), and so on.

Actually many finer details are involved. But it is remarkable that even gross approximations as given here are able to explain the general nature of X-ray spectra quite well.

A point to note is that although $K\alpha$ line corresponds to energy 65·5 kev, creation of a vacancy in K-shall would require 82·7 kev energy. Hence an X-ray tube run at voltage smaller than $\sim 83\ kV$ would not give K-lines with gold target. This has been actually verified.

Moseley in 1913 announced that *the frequencies of corresponding X-ray lines in different elements are approximately proportional to square of the atomic numbers of the elements.* This is called Moseley's law. With more detailed observations the law is now stated as

$$\nu = KR(Z-\sigma)^2. \qquad ...(20{\cdot}18)$$

where σ is called the screening constant and has different values for $K, L, M, ...$ series. R is Rydberg constant and K is a constant for each series.

On the basis of approximate effective central charges $(Z-1)$ and $(Z-9)$ used by us for the K- and L-shell electrons, an estimate for the $K\alpha$ line would be :

$$\bar{\nu} = R(Z-1)^2 - R(Z-9)^2\frac{1}{2^2}$$

$$\approx R(Z-1)^2\left(\frac{1}{1^2}-\frac{1}{2^2}\right)$$

$$= \frac{3}{4}R\,(Z-1)^2 \qquad \text{...(20·19)}$$

This fits Eq. (20·18) with $\sigma=1$ and $K=3/4$. For the $L\alpha$ line our approximation would give $\sigma=9$:

$$\bar{\nu} = R(Z-9)^2\left(\frac{1}{2^2}-\frac{1}{3^2}\right) \qquad \text{...(20·20)}$$

while the experimental value is $\sigma=7·4$.

X-ray Absorption Edges. X-rays passed through a material may show (i) general scattering, (ii) Compton scattering, and (iii) characteristic absorption. The first occurs because the Z electrons of the atom oscillate under the X-ray field of frequency ν and hence radiate in all directions at the same frequency. This is called *coherent scattering* and its intensity leads to determination of Z. Presently our interest is not in absorption on this account. Compton scattering is an incoherent scattering (see Chapter 19) arising from interaction with a valence electron.

The characteristic absorption of X-rays is due to their inelastic interaction with the *inner shell electrons* of the atom. This absorption, however, does not show sharp frequencies like the K, L, M ...emission X-ray lines. The reason is this. Appearance of an absorption frequency at $K\alpha$ position would need the transfer of a K-electron to L shell ; but there is no vacancy at the L-shell to permit this. Similarly, for absorption at $L\alpha$ position an electron must pass from L shell to the M shell ; but there is no vacancy at the M shell to permit this. For this reason no absorption *lines* are observed at the frequencies of the corresponding emission lines.

However, a K or L or M electron can go outside the atom which will need a minimum energy of $E=Rhc\,(Z-\sigma)^2/n^2$, with appropriate σ and n values. Since energy states outside the atom are not quantized, the absorption corresponding to (say) K electron ejection is continuous beyond the frequency ν_K given by $h\nu_k=E_K$. It so happens that the maximum probability of such interaction occurs at $\nu=\nu_K$ and falls off for $\nu>\nu_K$. That gives an *absorption edge* at $\nu=\nu_K$. Similar statements apply for absorptions corresponding to the ejection of L, M...electrons.

It will be seen that *differences* of ν values of the absorption edges, give the ν values of the emission lines in X-ray spectra.

The Continuous X-ray Spectrum. The continuous (or white) X-ray emission spectrum is due to the retardation of the impinging electrons of the X-ray tube in the body of the solid anticathode in general, without ejecting an electron from any shell of an atom. The energy loss can be in any number of steps, of different sizes, and hence the emission can be of any frequency in general. But the extreme limit is that all the energy eV may be lost in one step. Then the frequency will be maximum, given by

$$h\nu_{max}=eV \quad \text{or} \quad \lambda_{min}=\frac{hc}{eV} \qquad \text{.. (20·21)}$$

where V is the accelerating voltage in the X-ray tube. The general spectrum will thus have an *upper frequency limit* or lower wavelength limit. This is called the *Duane and Hunt* limit.

Example 20·8. Deduce the Duane and Hunt limit for an X-ray tube working at 40 kV.

Solution. Using eq. (20·21)

$$\nu_{max}=\frac{40000\times1\cdot6\times10^{-12}}{6\cdot62\times10^{-27}}\ Hz=0\cdot97\times10^{19}\ Hz$$

$$\lambda_{min}=\frac{c}{\nu_{max}}=0\cdot31\times10^{-8}\ \text{cm.}$$

PROBLEMS

20·1. Deduce the potential due a nucleus of $Z=92$ at a distance 4×10^{-14} m. Compute the *minimum* kinetic energy of a (*i*) proton, (*ii*) deuteron, (*iii*) α-particle so that it can reach the nucleus upto this distance.

20·2. Consider a silver foil 2×10^{-5} cm thick. Density of silver being 10·5 g. cm^{-3}, deduce (*i*) the number of atomic layers in the foil, (*ii*) the number of silver atoms per cm^2 in each layer, and (*iii*) the probability that an incident α-particle passes within $1\cdot0\times10^{-14}$ m from the centre of a nucleus.

20·3. Discuss how the Rutherford scattering experiment and its conclusions would be affected if any one of the following changes were made : (*i*) the foil is taken 1000 times thicker, (*ii*) the foil is taken 10 times thinner, (*iii*) slower α-particles are taken, (*iv*) neutrons are used as projectiles.

20·4. In a Rutherford experiment α-particles of energy upto 3·2 *Mev* suffer 180° scattering with a gold foil. Deduce your conclusion from this, starting any assumptions involved.

20·5. Discuss the conclusions drawn from the angular distribution of scattered α-particles. Examine the statement that occurrence of back-scattering cannot by itself settle the size of the nucleus.

20·6. Consider a model in which charges $+80e$, $-20e$ and $-60e$ reside on concentric spheres of radii 4×10^{-12} cm, 8×10^{-12} cm and 12×10^{-12} cm respectively. Check this model against the statements that protons of energy (*i*) 1 *Mev*, (*ii*) 2 *Mev*, (*iii*) 5 *Mev* suffer back scattering from this atom.

20·7. Describe Rutherford model of the hydrogen atom, and deduce the energy in electron-volt when radius of the electron orbit is 10 angstrom. What is the significance of the *minus* sign in this energy ? In what respects does this model fail ?

20·8. "Some of Bohr's postulates are in flagrant violation of the accepted laws of electromagnetism." Discuss critically.

20·9. Given Rydberg constant $R=109678$ cm^{-1}, deduce the energy level diagram of atomic hydrogen and show the lowest five energy levels to scale. Show the transitions corresponding to (*i*) the Balmer lines in emission, (*ii*) the Lyman lines in absorption.

20·10. For a particular atom the ionisation potential is 10·4 volt. Calculate the wavelength of the principal series limit.

20·11. Deduce the wavelengths corresponding to the first two absorption lines and the ionisation limit for He^+ atom, given that isolation limit for H atom is at $\lambda=900A$.

20·12. For He^+ atom calculate (*i*) radius of the first orbit, (*ii*) frequency of orbiting, and (*iii*) centripetal acceleration.

20·13. Name the physical property governed by the quantum number (i) n, (ii) l, and (iii) m_l. State all the possible values of l for $n=2$, and all the possible values of m_l for $l=1$.

20·14. Show that for a completed sub-shell in atomic structure $\Sigma m_l=0$ and also $\Sigma m_s=0$. What physical quantities consequently become zero ?

20·15. State Pauli's exclusion principle. Show that the maximum number of electrons in a state of total quantum number n will be $2n^2$.

20·16. The full K, L, M, N shells can contain 2, 8, 18, 32, electrons respectively. But atom $Z=19$ has the distribution 2, 8, 8, 1. Write the electronic configuration, and comment why the last electron goes to N shell though M shell is not full.

20·17. The energies of ns orbits in an atom are given by $-R/(n-1{\cdot}80)^2$ and those of $n'p$ orbits by $-R/(n'-0{\cdot}91)^2$, where n and n' are total quantum numbers. Draw the energy level diagram for $n=3, 4, 5, \infty$ and $n'=3, 4, 5, \infty$. Deduce the wavelength for transition $3s\rightarrow4p$.

20·18. With the data of Q. 20·17, compute the first three excitation potentials and the ionization potential, given lowest n and $n'=3$.

20·19. Explain the significance of the following :

(*a*) core electrons, (*b*) valence electrons, (*c*) penetrating orbits, (*d*) quantum defect.

20·20. Discuss the origin of X-ray emission spectra. State Moseley's law and give an elementary explanation for it.

20·21. Compute the radii of the K-shell and L-shell in an atom of $Z=30$, assuming effective shielding as that of 1 and 8 electrons respectively. Compare them with Bohr radius a_0.

20·22. Estimate in electron volts the energy required to eject an electron from (*i*) the K-shell, (*ii*) the L-shell of an atom with $Z=92$, given that ionisation potential of H-atom is 13·6 *ev*. Hence deduce the estimated wavelength of $K\alpha$ line of this atom.

20·23. An X-ray tube is run at 25000 volts. Deduce the short wavelength limit of the continuous X-rays.

20·24. The K-, L- and M- absorption edges for gold fall at 0·153A, 0·902A and 3·61 A respectively. Deduce the wavelengths for $K\alpha$, $K\beta$ and $L\alpha$ lines.

20·25. From Eq. (20·9), $E=-Ze^2/2r_n$, one may use Ritz Principle and experimental value of R to set $E=-Rhc/n^2$ for hydrogen ($Z=1$). Hence deduce expressions for v_n, r_n, and mv_nr_n. From known values of consistants, calculate the value of mv_nr_n $n=1$. See if you find that this value is related to h, showing that R is not a fundamental but a derived constant.

21

Coherence and Lasers

In Chapter 11 we defined *coherence betweeen two sources* of light as the existence of a constant phase relation between them. We now know that what we observe as light from a source comes as the sum total of radiations by billions and billions of atoms (or molecules) in the source; also that an individual atom too does not radiate continuously, but only occasionally when it undergoes a quantum transition downward after initial excitation. The question therefore arises :

To what extent may the radiations from different atoms of a given source be related in phase, in direction of emission and in polarization ?

This question, in effect, speaks of the *coherence of a given source*. This will be discussed in this Chapter.

The important fact is that in recent years we have developed sources which are highly coherent, *i.e.*, those in which the radiations by almost all the emitters in the source are in mutual agreement not only of phase, but also of the direction of emission and the polarization. These *coherent sources* are called *lasers* and they are capable of a very wide range of applications in science, industry, technology, communications, medicine, etc.

We shall first discuss the coherence of ordinary sources and then consider the development of lasers.

21.1 Coherence : Purity of a Spectral Line. In § 9.6 we had discussed wave-groups. All wave-trains have a finite length. Earlier in § 9.2 we had discussed the quality factor Q of an oscillator, from which wave-trains must arise. A wave-train like that of Fig. 9.6 (*b*) or 9·5 (*a*), treated by the Fourier's transform method shows that it contains waves of frequencies spread on both sides of ω_0, the width $\Delta\omega_{1/2}$ [Fig. 9.5 (*b*)] being in inverse proportion to length b of the wave-train. Let us now shift attention to the *experimental* observations.

A *spectroscope* shows that each spectral line of wavelength λ has a finite spread. The cadmium red line ($\lambda = 6438\ A°$) has a spread $\sim 0{\cdot}007\ A°$; sodium yellow D_1 line ($\lambda = 5896\ A°$) has a spread $\sim 0{\cdot}05\ A°$. These spreads are *not* due to the observing instruments, because instruments of resolving powers capable of showing $\Delta\lambda$ less than 0·001 A° were used. So we see *experimentally* that each spectral line has a finite purity. We may *define this purity* Q by*

$$Q \approx \frac{\lambda}{\Delta\lambda} \qquad \text{...(21.1)}$$

where $\Delta\lambda$ is half-width of the spectral line. We then see that the purity of *Cd* red line is $\sim 10^6$ and of *Na* D_1 line $\sim 10^5$.

A *Michelson interferometer* experiment with light of a "single" spectral line shows that if the difference of the two paths is increased continuously, the dark-and-bright contrast in the interference at fringes falls off, till the fringes vanish. The path difference at which this occurs is about 10-fold larger for *Cd* red than for *Na* D_1. The observed path difference at which this occurs is called the *coherence length* of the concerned spectral line, denoted by *L*.

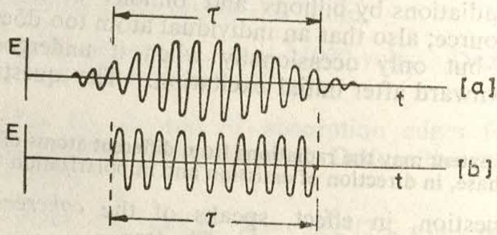


Fig. 21·1.

Two ways of showing emission by an atom or molecule

The models of Fig. 21·1 explain these. Consider [Fig. 21.1 (*b*) which says that a given emission lasts only for an average time τ, called the *coherence time*. In a Michelson interferometer the wave-train takes two alternative paths ; each train has spread τ. A path difference *d* between the two routes means time difference d/c. If this exceeds τ then the waves coming from the two paths would no more have a phase relation and so the interference fringes would disappear. Since the disappearance of fringes has been observed at $d=L$ we get $L/c \approx \tau$, or

$$L \approx c\tau \qquad \text{...(21·2)}$$

Next consider Fig. 21·1 (*a*). If the radiation is assumed to have a frequency spread $\Delta\nu$ around the mean frequency ν, then the fading at the tails is broadly explained by setting that the number of oscillations of frequencies ν and $\nu+\Delta\nu$ in time τ differ by unity.⁺

---

* This quantity is related with what is called the quality factor *Q* of a classical oscillator. (See § 9·2)

⁺ We are not dealing with just two waves of frequencies ν and $\nu+\Delta\nu$ but with a continuous set in the range $\Delta\nu$. Hence we use here unity, not $\frac{1}{2}$. In any case, these are just order-of-magnitude statements. Rigorous procedure is the Fourier transform method § (9·3.)

This means

$$\tau(\nu+\Delta\nu)-\tau(\nu)=1 \Rightarrow \tau\Delta\nu=1 \qquad \text{...(21.3 a)}$$

But $|\Delta\nu|=c\Delta\lambda/\lambda^2$. Hence

$$\Delta\lambda=\frac{\lambda^2}{c\tau}=\frac{\lambda^2}{L} \qquad \text{...(21.3 b)}$$

A Fourier transform of the curve of Fig. 21·1 (*b*) would also give the same result. Thus the coherence length *L* (Michelson interferometer expt.) and spectral line width $\Delta\lambda$ (spectroscopic expt.) are both related with the coherence time τ (model of Fig. 21·1). In terms of purity Q of the spectral line (Eq. 21·1).

$$c\tau \approx L \approx \lambda Q \qquad \text{...(21·4)}$$

Thus coherence time τ, coherence length *L* and purity *Q* are three equivalent ways of expressing the coherence of the observed light.

Example 21·1. The coherence length for sodium D_2 line is 2·5 cm. Deduce (i) the coherence time τ (ii) the spectral width of the line, (iii) the purity factor Q.

Solution. From Eq. (21·2)

$$\tau\approx\frac{L}{c}=\frac{2·5 \text{ cm}}{3·0\times10^{10}\text{ cm/s}}\approx0·8\times10^{-10}\text{ s}$$

From Eq. (21·3 *b*), taking $\lambda\approx6\times10^{-5}$ cm

$$\Delta\lambda\approx\frac{\lambda^2}{L}=\frac{36\times10^{-10}\text{ cm}^2}{2.5\text{ cm}}\approx14\times10^{-10}\text{ cm}\approx0·14A$$

From Eq. (21·1)

$$Q=\frac{\lambda}{\Delta\lambda}\approx\frac{6\times10^{-5}\text{ cm}}{14\times10^{-10}\text{ cm}}\approx0·4\times10^{5}$$

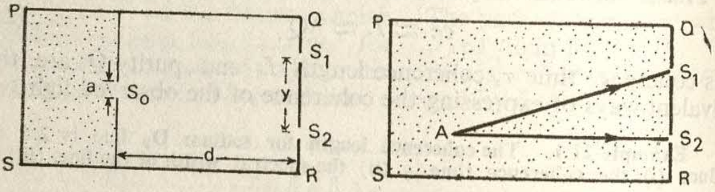
Spatial Coherence Length. The coherence length of Eq. (21.4) is the distance *along the line of sight* over which phase-relation is expected to hold. Over what distance *transverse* to the line of sight do we expect the source to be coherent, *i.e.*, showing definite phase relations? In Fig. 21.2, *PQRS* is the source, and the question here can be answered by having two slits S_1 and S_2 at separation y and looking for the interference pattern outside on the right. If interference occurs for $y_{max}=L_w$, then this dimension is called the *spatial coherence length* of the source *PQRS*. If the dots represent the excited atoms, then in case (*a*) the source is really a distance *d* away from S_1 and S_2, with a slit S_o at its front. In chapter 14 we have seen that if slit S_0 has width *a*, the light from each emitter atom would spread to an angle $\sim\lambda/a$. Hence if angle $S_1S_0S_2$ $(=y/d)$ is less than λ/a then both S_1 and S_2 will receive light from *each* of the emitters and will thus be in coherence. Thus the condition for coherence of S_1 and S_2 becomes

$$\frac{\lambda}{a}>\frac{y}{d} \qquad \text{or} \qquad y<\frac{\lambda}{a/d}$$

The maximum value of y is called L_w. Also a/d is the angle (θ) subtended by the width of slit S_0 at the plane of observation. So we get

$$L_w = \frac{\lambda}{\theta} \qquad \ldots(21{\cdot}5)$$

Thus in this case we can control L_w by adjusting a and d. How about the source $PQRS$ of Fig. 21·2 (b) ? Here S_1 and S_2 are directly on the source. The experimental fact is that in this case interference fringes have not been observed for separation $S_1 S_2$ as small as 10^{-8} cm. Reason is this ; For an emitting atom A, the path difference $AS_1 - AS_2$ will govern the phase relation at S_1 and S_2. But with A varying in space, the phase relation at S_1 and S_2 will also vary. So no constant phase relation exists between S_1 and S_2. We say that in this case source $PQRS$ has very little spatial coherence.



(a) (b)

Fig. 21·2.

Two 'sources' PQRS with slits S_1 and S_2 for discussing coherence between the emergent beams

Laser* is a device to increase both L and L_w. Increasing L amounts to making the successive wavetrains like the one in Fig. 21.1 (b) fit in phase, and increasing L_w amounts to ordering atoms perpendicular to the line of propagation to emit in the same phase. We will see how this is achieved.

21.2 Stimulated or Induced Emission. In Fig 21.3 we show two energy states 1 and 2 of an atomic system. The black dots schematically show the number of atoms in each state, and three different events are shown at (a), (b) and (c). In each case the situation *before* and *after* an event is shown. The events may be described as follows :

(a) **Spontaneous emission.** One of the atoms passes *spontaneously* from state 2 to state 1, emitting a photon $h\nu_{12}$.

(b) **Stimulated or induced emission.** A photon of frequency ν_{12} from the surrounding radiations *induces* or *stimulates* one of the atoms from state 2 to go to state 1. So two photons $h\nu_{12}$ come out.

---

*'Laser' is an acronym for Light Amplification by Stimulated Emissiom of Radiation. See case (b) in Fig. 21.3.

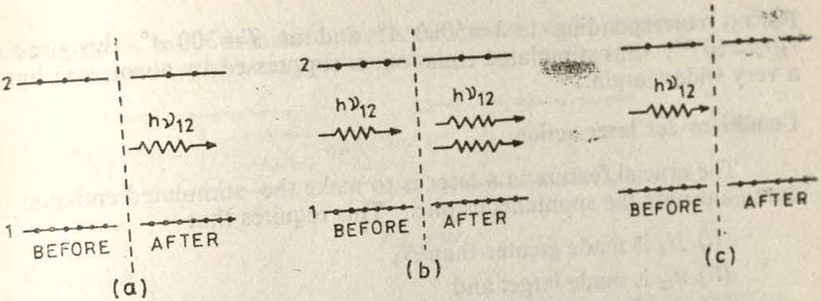


Fig. 21.3.

Three possible optical events in a 2-level system : (a) spontaneous emission, (b) stimulated emission, (c) absorption.

(c) Absorption. One of the atoms passes from state 1 to state 2 under stimulation from a photon $h\nu_{12}$ of the surrounding radiation, so that photon is lost.

Einstein's A and B Coefficients.

Now, events (*a*) and *c*) were known before the year 1917. In that year Einstein suggested that the stimulated transition 2→1 should also be possible exactly the same way as 1→2 is. He discussed the relative probabilities of the the three events of fig. 21.3 in terms of the populations N_1 and N_2 of atoms in the states 1 and 2 respectively, and the radiation density u_{12} of the radiation of frequency ν_{12} present in the region of interest. The rates at which they occur are—

$$\left.\begin{array}{ll} r_a \text{ (spontaneous emission)} & = A_{12}N_2 \\ r_b \text{ (stimulated emission)} & = B_{12}N_2u_{12} \\ r_c \text{ (absorption)} & = B_{12}N_1u_{12} \end{array}\right\} \qquad \ldots(21.6)$$

The coefficients A_{12} and B_{12} used here are called *Einstein's A and B coefficients.* The important fact is that the same coefficient B_{12} determines the rates for stimulated emission and absorption.*

The most important fact about the *stimulated* emission is that the photon thus produced is coherent with the stimulating photon—'Coherence' here means more than agreement of phase. Here agreement in *direction* of travel and agreement in *polarization* are also included.

However, in ordinary light the spontaneous emission dominates. The reason is that Eq. (21.6) and Boltzmann statistical law give, of temperature T,

$$\frac{r_b}{r_c} = \frac{N_2}{N_1} = e^{-h\nu_{12}/kT} \qquad \ldots(21.7)$$

---

*For this reason stimulated emission is also called *inverted absorption.*

For ν_{12} corresponding to $\lambda=5000\ A^\circ$ and at $T=300\ A^\circ$, this gives $r_b/r_c \approx 10^{-44}$, thus stimulated emission is suppressed by absorption by a very wide margin.†

Conditions for laser action

The crucial feature in a laser is to make the stimulated emission dominant over the spontaneous one. This requires that—

(*i*) N_2 is made greater than N_1
(*ii*) u_{12} is made large; and
(*iii*) A_{12}/B_{12} is kept small.

Making $N_2 > N_1$ is called *population inversion*, because it is against the equilibrium condition.‡

Making u_{12} large is achieved by enclosing the emitted radiations in a *cavity*, usually between two parallel reflectors. The radiation repeatedly travels back and forth, only a small part being able to go out. In this cavity large u_{12} is thus created, and stimulated emission grows in strength relative to the spontaneous one.

Small A_{12}/B_{12} is achieved by choosing the upper level 2 as a *metastable* level, *i.e.*, one from which spontaneous transitions to level 1 are normally disallowed, making A_{12} extremely small.

It may be noted that through the basic concept of stimulated emission was given by Einstein in 1917, the first suggestion for using it for a specially coherent source was given in 1954 by Townes and the first device was produced in 1955 by Gordon.

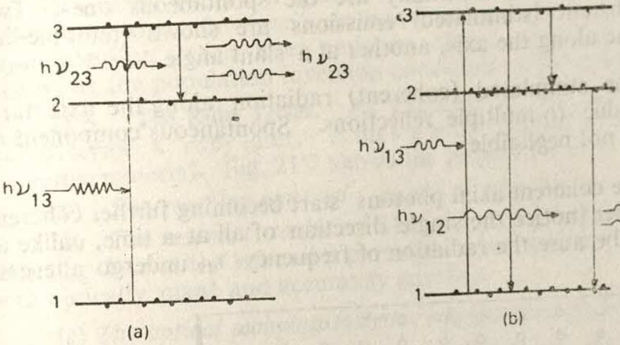
21.3. **Pumping and Population Inversion.** A system in which population inversion is to be achieved is called an *active system* (for laser). Raising the molecules to the higher energy state is called *pumping* and this can done in many ways. Ammonia gas was the first active system used. It has two states with level separation corresponding to $\lambda=1.27$ cm (microwave region). Ammonia gas is heated and then passed through a non-uniform electric field, which deflects the excited and unexcited molecules to different regions, thus causing $N_2 > N_1$ in one region.

Optical Pumping is a more common method. In fig. 21.4 we describe two 3-level systems. At (*a*) is a system in which pumping by light of frequency ν_{13} raises the atoms to state 3, from where they may go back to states 1 or 2. The conditions $N_3 > N_2$ is created by this pumping, and for light of frequency ν_{23} the stimulated emission in this system may dominate over absorption; with large u_{23} it will also dominate over spontaneous emission.

---

†Examine why the first success in making stimulated radiation dominent was achieved in the microwave region.

‡Eq. (21.7) would show that $N_2 > N_1$ corresponds to negative T. Hence achieving 'population inversion' is also called achieving an effective 'negative temperature' state.



(a) (b)

Fig. 21.4.

Producing inverted population in a 3-level system
(a) Making $N_3>N_2$, (b) making $N_2>N_1$

In Fig. 21.4(*b*), is a system in which state 2 is a *metastable state*, which really means that the spontaneous emission coefficient A_{12} is very low. Optical pumping raises the atoms to level 3. From there most of the excited atoms return to state 1, but some go to state 2. Since transition 2→1 has very low probability, in due course of time the population N_2 exceeds N_1 if pumping power is sufficient to keep level 2 fed at a faster level then its depletion by 2→1 transition. (See also § 21·5).

Pumping by electrical discharge and by some other methods has also been used in several cases. But we shall describe them later (§ 21.6).

21.4. Cavity Resonators and Shaping of a Beam. In a laser the active system is enclosed in an optical cavity, usually made of two parallel surfaces, one of which is a good reflector and the other with ~90% reflectivity (or 10% transmission). The cavity makes the energy density large*, whence stimulated emission dominates over the spontaneous emission. The multiple reflection also makes the stimulated emission more coherent—a process called *shaping* of the beam.

Fig. 21.5 schematically shows how the beam shapes, with time. It may only be taken as a schematic representation, showing the dominant fact at each stage. The hollow circles and black dots here represent the unexcited and excited atoms respectively. The left end mirror represents the full reflector and the right end one the partial reflector. The description is as follows :

(*a*) Almost all atoms are initially in the unexcited state.

(*b*) When the pumping radiation falls on the active material the atoms start reaching the excited state. (Population inversion not yet reached, black dots less then hollow circles).

---

*See Problem 21·5.

(*c*) Pumping radiation continues (but not shown in this and successive figures). Population inversion is reached at this stage. The dominant radiations initially are the spontaneous ones. Two cases of coherent (stimulated) emissions are shown (multiple-line arrows)— one along the axis, another at a slant angle.

(*d*) The stimulated (coherent) radiation along the axis starts dominating due to multiple reflections. Spontaneous component of emission still not negligible.

(*e*) The coherent axial photons start becoming further coherent with each other (notice the single direction of all at a time, unlike at (*d*). This is because the radiation of frequency ν_{12} undergo alternate absortptions/emissions and those that are mutually coherent go on becoming dominant at successive stages. Spontaneous component of emission is now very weak.

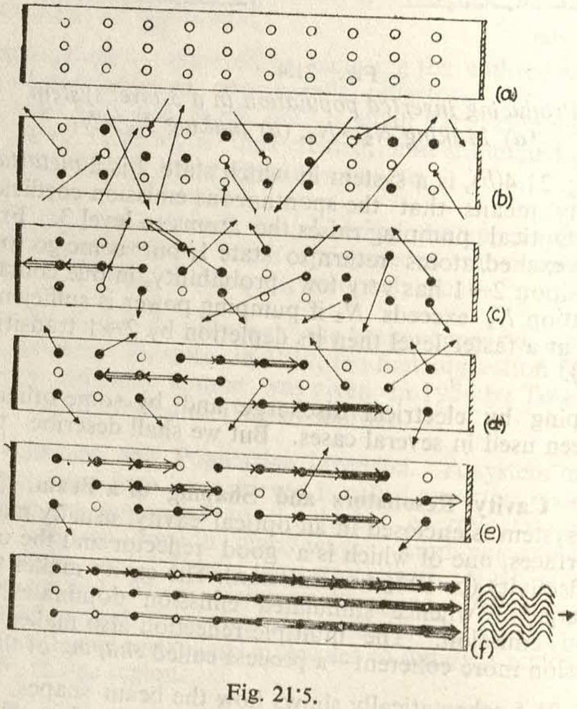


Fig. 21·5.

Shaping of the coherent beam in the laser cavity resonator (see text)

(*f*) Almost the entire emission becomes coherent. The wave-shapes shown at the exit on the right show phase agreement of waves even from laterally separated atoms.

In a *continuous wave laser*, *stage* (*f*) may be treated as a steady state in which the pumping rate equals the rate at which spontaneous plus stimulated radiation comes out of the laser. But there are *pulsed lasers* also. In these a strong pumping pulse creates the population inversion ; then laser emission starts, shapes and grows, and falls off as the population inversion condition no more holds.

21·5. The Ruby Laser. As example of an actual laser we will describe a ruby laser. It uses a solid crystalline rod of ruby as the active material. Fig. 21·7 shows the general features :

(*a*) *The working material* : a rod of ruby crystal.

(*b*) *The resonant cavity* : The ruby rod itself has a fully reflecting face at the left and a partially reflecting one at the right, both optically plane and accurately parallel.

(*c*) *The optical pumping system* : a helical xenon discharge tube.

(*d*) *The cooling system* : water circulation in a glass tube surrounding.

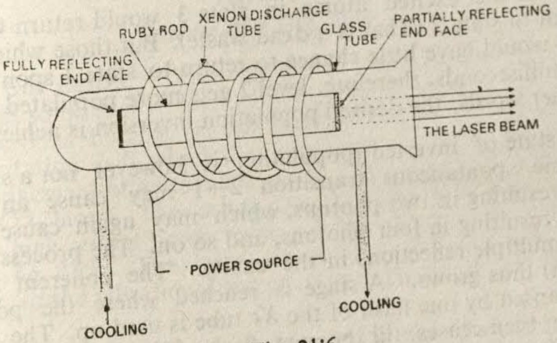


Fig. 21·6.

The general features of a ruby laser

Ruby is crystalline Al_2O_3 with some Al atoms replaced by *Cr*. The active material in the ruby is Cr^{3+} ions giving it a pale pink colour at ~0·055% concentration of chromium. A typical rod may be 10 cm long and 0·8 cm in diameter. The reflectors at the ends may be direct coatings on the ruby rod itself or separate pieces attached to it. For the partially reflecting end the transmission coefficient is ~10%.

In the xenon flash tube each flash lasts several milliseconds and in each flash a few thousand joules of energy is consumed. Since the xenon discharge gives white radiations and ruby is a 3-level system (see fig. 21·7), only a small part of the energy is used in pumping, *i.e.*, in exciting the Cr^{3+} ions. The rest goes to heat up the apparatus ; hence the need for cooling by water circulation.

In ruby the 'pumping' corresponds to ~5600 $A°$ (green region), and the absorption band has a width nearly 800 $A°$* (Fig. 21·7). The transitions 3→2 are non-radiative, *i.e.*, the Cr^{3+} ions give the corresponding energy to the crystal lattice (in the form of heat). The relative probabilities for the three downward transitions may be expressed as follows :

---

* If the band were narrower, a smaller part of the *Xe* flash energy would be useful for pumping.

$$p_{3\to1} \approx 300\, p_{3\to2} \approx 6\times10^{4}\, p_{2\to1}$$

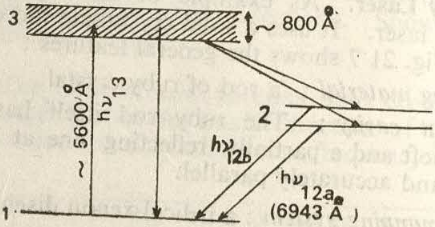


Fig. 21·7.
Energy levels of ruby

Thus most of the excited atoms in state 3 would return to state 1, rather than to state 2 (That is a dead waste). But those which go over to state 2 would have little chance to return to state 1 spontaneously. In a few milliseconds, therefore, level 2 gets more populated than level 1. In other words, the desired population inversion is achieved.

The state of inverted population is however not a stable one, because one spontaneous transition 2→1 may cause an induced transition resulting in two photons, which may again cause induced transitions resulting in four photons, and so on. The process goes on, helped by multiple reflections in the cavity. The coherent radiation (laser beam) thus grows. A stage is reached where the population inversion caused by one flash of the *Xe* tube is used up. The coherent (laser) beam then ceases, till the next flash of the *Xe* tube repeats the process. Thus the ruby laser is a *pulsed laser*.* It means that the active system (here Cr^{3+} ions) is excited by the *Xe* helical discharge in pulses, and the system emits laser light in pulses. While the *Xe* pulse is of several millisecond duration, the laser pulse is of much less than a millisecond duration—which means enhanced instantaneous power.

A small detail needs mention. Level 2 in ruby is really a doublet of separation 14*A*°. The spontaneous transitions would give two weak lines at 6943 and 6929 *A*°, each of width ~6*A*°. But under lasing conditions the line 6943 *A*° from level 2*a* dominates by a very wide margin and its spectral width narrows down by over 100-fold.

21·6 Various Kinds of Lasers. A very large range of lasers have been developed—both of the pulsed and continuous wave type—using various active materials, various pumping devices, and various resonant cavities. Output powers as high as 100 kilo-watt and *Q* values $\sim 10^{14}$ have been achieved. We shall here make only a broad mention of some features.

---

* Continuous wave lasers are also in use, including those of ruby. But we shall not describe them here.

(*a*) *Glass Lasers*. The ruby laser is one among many glass lasers. Here Al_2O_3 was the matrix and Cr^{3+} the activator. Among other combinations used are CaF_2 with U^{2+}, Dy^{2+} or Nd^{3+} and SrF_2 with U^{3+}, Sm^{2+} or Tm^{3+}.

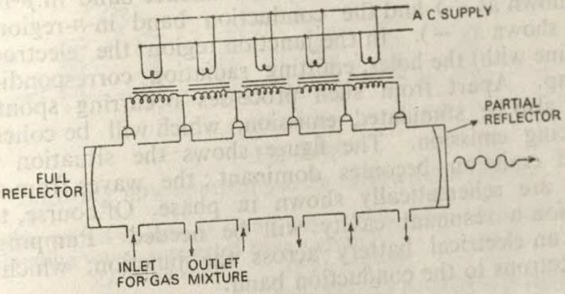


Fig. 21·8.

Scheme of a CO_2 laser

(*b*) *Gas Lasers*. These require a mixture of two gases *A* and *B* such that some excited level of A falls close enough to an excited level of *B*. Gas dischage in the mixture excites *A*, collisions transfer the excitation to *B*, which then produces the laser beam with the usual arrangement of a resonant cavity. Typical examples are $He-Ne$ and $N_2—CO_2$ lasers. The carbon dioxide lasers have been developed upto very high output power levels. A typical set-up for CO_2 laser is shown in fig. 21·8. The laser emission is like that in fig. 21·4 (*a*), belonging to the vibrational transition in CO_2, which falls in infrared region.

(*c*) *Dye Lasers*. These employ organic dyes or inorganic liquids as the active media, and the excitation is done by optical pumping. Dyes give the possibility of generating a laser beam at any frequency in the infrared and visible regions and laser excitation efficiency reaches as high as ~50%. Typical dyes used for the visible range are rhodamines and pyronine ; the solvents may be water, alcohol or glycerol.

A special feature of dye lasers is that *tunable lasers* can be made from them—*i.e.*, λ variable over a wide range and still keeping the characteristic of laser light. The method is this ; another laser falling on the dye gives fluorescence over a broad range of λ. Now, the cavity involves a grating monochromator, which selects a particular λ out of the fluorescenet radiation. The laser action occurs at this λ and turning the grating angle provides the tuning, *i.e.*, variation of chosen λ over a desired angle.

Glass lasers (like ruby laser) have the advantage of high concentration of active particles (like Cr^{3+}), but suffer from the technological difficulty in getting high optical homogeneity of the medium. On the other hand gas lasers can be had to dimensions as large as several tens of metres length and 20 cm diameter, but have low concentration of the active particles per unit volume (operation at usually quite low pressures). The liquid lasers (including dye ones) combine the advantages of both the glass and lasers. But they are suitable for use as wide-band amplifiers of light, not where highly monochromatic radiation is needed.

(*d*) *Semiconductor Lasers.* Fig. 21·9 shows the scheme of an *n-p* junction in a semi-conductor. The valence band in *p*-region has 'holes' (shown as +) and the conduction band in *n*-region has free electrons shown as −). In the junction region the electrons fall in (or combine with) the holes, emitting radiation corresponding to the energy gap. Apart from such processes occurring spontaneously, there will also be stimulated emission, which will be coherent with the inducing emission. The figure shows the situation when the stimulated emission becomes dominant ; the waves from successive emissions are schematically shown in phase. Of course, to achieve this situation a resonant cavity will be needed. Pumping is done simply by an electrical battery across the junction, which goes on feeding electrons to the conduction band.

Since there is a finite range of energy over which the electron-hole combination occurs, these semiconductors give wide band amplification. The light is usually in the infrared region.

The current density needed to initiate lasing action is a few thousand amp./cm^2, and the emitted beam becomes spectrally narrower (more coherent) as the current density increases.

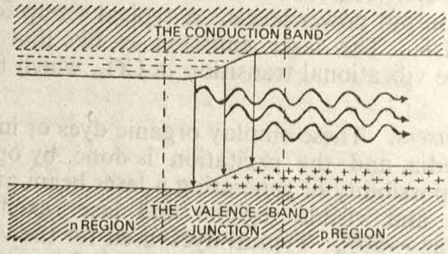


Fig. 21·9.
Energy levels for a Semiconductor Laser

21·7. High Monochromaticity of Laser Radiation. We have stated that the photon emitted by an excited atom under stimulation by another photon is coherent in phase. In terms of Fig. 21·1 (*b*) this would mean that the waveforms from stimulated emissions by different atoms chain-up *in phase*, so as to form a continuing sinewave character over a much larger time τ than for spontaneous emission. In Fig. 21·1 (*b*), as also in Example 21·1, we have $\tau \approx 10^{-10} s$. This is for spontaneous emissions. For a laser beam $\tau \approx 10^{-6} s$ or even more ; about 10^4 fold increase of Q, that is, spectral line width $\Delta\lambda_{1/2}$ smaller by that factor.

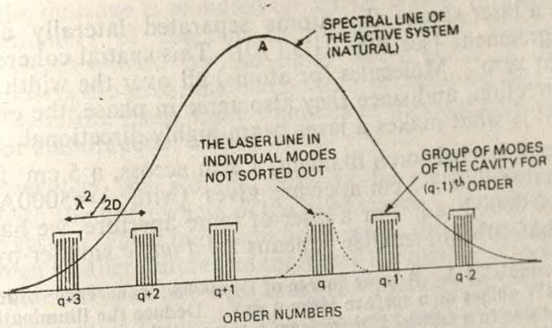

SPECTRAL LINE OF THE ACTIVE SYSTEM (NATURAL)
A
THE LASER LINE IN INDIVIDUAL MODES NOT SORTED OUT
GROUP OF MODES OF THE CAVITY FOR (q-1)th ORDER
λ²/2D
q+3
q+2
q+1
q
q-1
q-2
ORDER NUMBERS

Fig. 21·10

Showing how spectral purity iucreases in a laser

Another way is to look at it this way. With mirror separation D in the cavity the standing waves supported have λ given by

$$2D\cos\theta = q\lambda. \quad (q \text{ an integer}) \qquad \ldots(21{\cdot}8)$$

The separation $\triangle\lambda_0$ for successive q values differing by unity is given by

$$\frac{\lambda}{\triangle\lambda_0} = \frac{2D}{\lambda} \qquad \ldots(21{\cdot}9)$$

For $D \sim 1$ m and $\lambda \sim 5\times10^{-7}$ m you thus get purity $Q \sim 4\times10^6$. Actually, depending on θ, each mode q splits into several discrete components called modes. Fig. 21·10 describes the situation. Within the spectral width of the active system some 5 or 6 q values occur, and for each q there are several modes at much smaller separation. The cavity dominantly encourages the q group nearest to the peak A at the cost of others. The dotted curve represents the profile of the laser beam achieved. With special design one can sort out an individual mode from a group : that is what leads to a million-fold purity.

Example 21·2. With a *He-Ne* laser, Michelson interferometer fringes remained clearly visible when the path difference was increased upto 8 m. Deduce the lower limits for (*i*) the coherence length, (*ii*) coherence time, (*iii*) spectral half-width, and (*iv*) Q of the line, (Given $\lambda = 11{\cdot}5\times10^{-5}$ cm.)

Solution. The experimental fact of observing interference for path difference upto 8 m directly gives (vide § 21·1)—

$$\text{coherence length } L > 8 \text{ m}$$

Eqs. (21·2), (21·3) and (21·4) then give—

$$\text{coherence time } \tau = \frac{L}{c} > 2{\cdot}7\times10^{-8} \text{ sec.}$$

$$\text{Spectral half-width } \triangle\lambda = \frac{\lambda^2}{L} \approx 1{\cdot}6\times10^{-3} \text{ A}^\circ.$$

$$Q = \frac{L}{\lambda} \approx 7\times10^6$$

21·8. High Directionality of Laser Radiation. In the usual sources of light a directional beam is obtained by collimation. This beam has two contributions to the angular spread : θ_1 due to diffraction at the aperture of the collimating lens, and θ_2 which equals the angle subtended by the source size at the centre of the collimating lens. Invariably, we have $\theta_2 \gg \theta_1$.

In a laser source the atoms separated laterally also emit in phase agreement [see Fig. 21·5 (*f*)]. This spatial coherence reduces θ_2 to near zero. Molecules (or atoms) all over the width emit in the same direction, and since they also agree in phase, the only spread is θ_1. That is what makes a laser beam highly directional.

For a typical torch filament, 1 mm across, a 5 cm focal length of collimator with 2·5 cm aperture gives (with $\lambda=5000$A), $\theta_2=0{\cdot}02$ rad, $\theta_1=0{\cdot}00002$ rad. For a laser of same aperture, we have θ_1 alone. Notice that 10^3-fold smaller θ means *solid angle* smaller by 10^6-fold.

Example 21·3. A laser source of $\lambda=6000$A, coherence width 4 mm and power 10mW shines on a surface 100m a way. Deduce the illumination. Compare it with that due to a focussed beam from a torch of filament diameter 1 mm, power 10W, lens focal length 10 cm and aperture 4 cm dia.

Solution. For the laser all 10 mW power goes in a cone of semiangle $\theta\approx\lambda/a=6\times10^{-5}/0{\cdot}4=15\times10^{-5}$ rad. Hence areal spread at 10^4 cm distance (besides the inital area $\pi\times0{\cdot}2^2$ cm^2) is—

$$A=D^2\pi\theta^2=10^8\times3{\cdot}14\times(15\times10^{-5})^2\approx7\ \text{cm}^2$$

$$\text{Illumination}=\text{Power/Area}=10\text{mW}/7\ \text{cm}^2=1{\cdot}4\ \text{mW/cm}^2$$

For the torch the 10W power spreads isotropically, and the lens collects a fraction $\pi\ \ (2\ \text{cm})^2\div4\pi\ (10\ \text{cm})^2=1/100$. The light spreads over a cone of semiangle determined by the filament size÷focal length of the lens$=0{\cdot}1/10=1/100$ rad. Hence areal spread at distance 10^4 cm is

$$A'=D^2\pi\theta^2=10^8\times3{\cdot}14\times(10^{-2})^2=3{\cdot}1\times10^4\ \text{cm}^2$$

$$\text{Illumination}=\text{Power/Area}=\frac{1}{100}\times10\text{W}/3{\cdot}1\times10^4\ \text{cm}^2\approx0{\cdot}003\ \text{mW/cm}^2$$

Example 21·4. If the laser beam of Ex. 21·3 is focussed by a lens of focal length 10 cm, deduce (i) the radius, (ii) area and (iii) power density of the image.

Solution. Angular spread of the laser beam is $\sim\lambda/d$, due to diffraction alone. Hence

$$\theta=6\times10^{-5}/0{\cdot}4=1{\cdot}5\times10^{-4}\ \text{rad}$$

$$\therefore\ \text{radius of the image}=f\,\theta=1{\cdot}5\times10^{-3}\ \text{cm}$$

$$\text{area of the image}=\pi(1{\cdot}5\times10^{-3})^2=8\times10^{-6}\ \text{cm}^2$$

$$\text{power density}=\frac{10\times10^{-3}\ \text{W}}{8\times10^{-6}\ \text{cm}^2}\approx1{\cdot}2\ \text{KW/cm}^2$$

21·9. Uses of Lasers. As stated earlier, the light from a laser has two special features not available in light from ordinary sources :

(*i*) *Narrow band-width*—which is the same thing as high monochromaticity or high temporal coherence.

(*ii*) *Narrow angular spread*—which is the same thing as high directionality (hence high intensity) or large spatial coherence.

The first character gives rise to uses of lasers typically in *communication systems*, computers and chemistry. In microwave communication[*] the message is mounted on carrier waves through modulation. The number of channels of message which can be carried simultaneously is limited by the finite band width of the carrier waves. If lasers are used, the band-width becomes smaller by a large factor and therefore many more channels of message can be accommodated for simultaneous transmission.

In *computers* the storage capacity for information similarly improves with the narrowness of the band-width. In any information processing system which uses radiation signals the capacity to distinguish between smaller differences increases the quality of performance. Lasers provide this through high monochromaticity of their beam.

Spectroscopy is known to be a powerful tool in Chemistry for investigating the structure of molecules. *Raman spectroscopy* is one in which lasers have made so much impact that a separate branch named 'laser Raman spectroscopy' has grown rapidly. One reason is that because of directionality much more exciting radiation can be concentrated on the sample (see Example 21·4). As a consequence Raman spectrum can be obtained from much smaller samples and much faster too. What is even more significant is that there are some special nonlinear interactions[+] which arise under such high intensity excitation only and these lead to considerable added information.

Narrow angular spread of the laser beam makes it a very useful tool for communications with earth satellites, and rockets to the moon and other planets. Lasers would also make more accurate range-finders. As typical examples, the earth-moon distance has been measured with the use of lasers[†] and studies on continental drifts are made with lasers.

One corollary of narrow angular spread is that laser beams can be directed into thin optical fibres ; by repeated total internal reflections the beams may travel several kilometers in these fibres. Thus optical fibre communication has become possible, which may replace telephone communication through wires or even microwave communication. The infrared lasers are best for this, because optical fibres have least losses in that region.

Another corollary of narrow angular spread is that laser beams can be focussed on very small areas, typically $\sim 10^{-6}$ cm^2. Thus with a laser of only $\sim$10mW power you get an image of intensity $\sim 10^5$ watt/cm^2. These considerations lead to use of lasers in medicine and industry. The laser beam is used to bore narrow holes through or cut thick sheets of metal because the high *power density* in its image melts the metal where it falls, allowing little time for conduction of the heat to other regions. In medicine micro-surgery has become possible, because one harmful component can be destroyed without seriously damaging the neighbouring regions.

---

[*] It is just like radio communication, except that λ is much smaller$\sim$1 cm.

[+] Polarization of the medium proportional to the second and higher powers of electric field E in the laser beam is at the root of these. See § 21·11.

[†] By measuring the time interval for receiving the reflected beam.

We have mentioned only a few typical examples here because the actual range of use of lasers is expanding very fast and associated devices are also being developed to enhance the scope. However, two applications directly related with wave optics will be described, *viz.*, holography, and non-linear optics.

21·10. Holography

(a) Introduction. When we view any three-dimensional object or scene, the full perspective is appreciated only by our freedom to change the viewpoint (position) and the focus. An ordinary photograph loses both these, since it records the view from a single viewing point and a fixed focus.

Gabor in 1947 suggested a different photographic technique using *coherent light* and *wave front reconstruction*. Although this technique was used by him in electron microscopy, it was not successfully realised in photography, because lasers had not been developed, and spatially coherent beams could be had with ordinary light sources only by employing very narrow pin-holes [see Fig. 21·2 (a)].

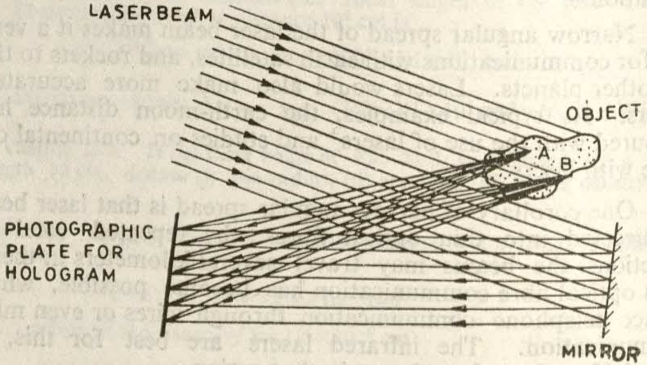


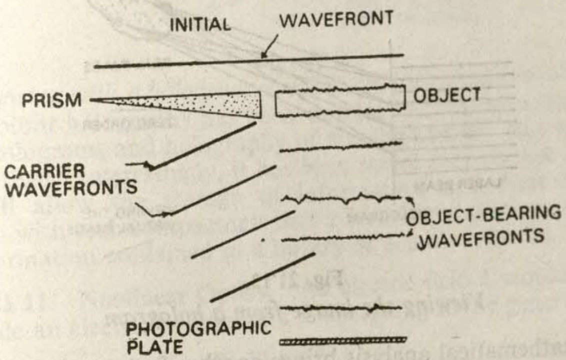
Fig. 21·11.
Obtaining the hologram of an object

With the advent of lasers we have now sources having both temporal coherence (monochromaticity) and spatial coherence. Leith and Upatnicks in 1962 revived Gabor's idea and since then there have been tremendous advances in what is now called *holography*. In Greek 'holos' stands for 'whole', and the records (*holograms*) of objects in the new technique contain details of both *amplitudes* and *phases* of light received from different parts of the three-dimensional object.

(b) Obtaining a hologram. Among several possible systems for making a hologram two are shown in figs. 21·11 and 21·12. The essence is to have the photographic plate illuminated simultaneously by (*i*) a carrier beam, and (*ii*) an object-bearing beam. Since both belong to the same laser wavefront the beams interfere on the plate. We may put, for the disturbance at any position x on the plate,

$$y=a_0 \cos(\omega t-kx)+a(x)\cos[\omega t-\phi(x)] \qquad \ldots(21\cdot10)$$

where the carrier (or reference) beam has amplitude a_0 and its phase is taken to vary *linearly* with x, and the object-bearing contribution at any point has amplitude a (x) and phase $\phi(x)$ coming from the sum-up of all the secondary wavelets from all the points of the object.*



Fia. 21·12.

Obtaining the hologram for a transparent object

This is in line with the Huygens-Kirchoff approach of § 10·1. It should be noted that relative phases of the wavelets which add up to the object-bearing beam depend on the angle of viewing and the depth differences.

Since the photographic plate records only the intensities, we get from eq. (21·10)

$$I=a_0^2+[a(x)]^2+2a_0\, a(x)\cos[kx-\phi(x)] \qquad \ldots(21\cdot11)$$

The significant point is that the intensity on the photographic plate has modulation involving both a(x) and $\phi(x)$. The plate when developed into a positive is called a *hologram* of the object. Its appearance has no similarly with the object at all, because each point on it has information about the *entire* object. The plate has only interference fringes which in their complexity store all the information needed to reconstruct in three dimensions the object holographed.

(c) **Viewing the Object.** For viewing the object from a hologram one has to use a laser beam of exactly the same wavelength as used for recording, and also the same geometry. Fig. 21·13 shows the arrangement corresponding to recording by the arrangement of fig. 21·11. The laser beam, passed through the hologram, has only the amplitude variations as dictated by eq. (21·11), but no information about phase variations. Therefore in the direct direction of the beam we have the zero order (like the diffraction grating), giving us no information. But at directions and distances corresponding to those in taking the hologram the wavelets from the hologram reconstruct the image in all detail, because information about $a(x)$ of Eq. (21·10) exists in the wavelets at start and information about ϕ (x) is generated the reverse way due to path differences.

---

* Fig. 21·11, for example, shows object point A contributing to all regions of the plate, so also B and every point on the object.

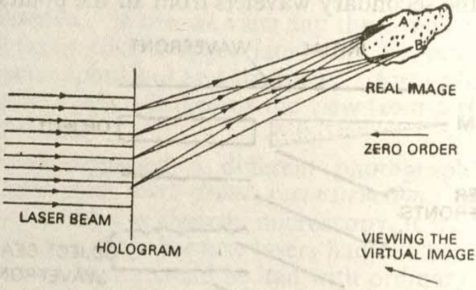


Fig. 21·13.
Viewing the image from a hologram

A mathematical analysis brings another fact : a virtual image is also seen at a position determined by the reflection of the real image in the plane of the hologram. This is not shown in fig. 21·13, but in practice it is this image which is more convenient to see.

The real image AB, or the virtual image (not shown) in Fig. 21·13 is *not* to be received on a screen. It is a three dimensional image to be *seen* from a position behind AB, and you may change the viewing position and the eye focussing just as you do in observing any three-dimensional scene.

(d) Some Special Features. Apart from the features already enumerated, one special feature of a hologram is that if the plate is broken into pieces ; each piece still would show the whole object. Only the quality of the image becomes poorer as smaller pieces are used in viewing.

Further, several images can be recorded in a single hologram (as many as 100, say) either by placing several objects in differently oriented locations and using a single reference beam for simultaneous recording, or by exposing the plate in succession, each time turning the hologram plate by an angle. Fig. 21·14 shows how from different eye positions E_1 and E_2 *the same part of the hologram* shows the images of different objects earlier holographed.

It is to be remembered that the depth and sideways width of a scene which can be recorded on a holograph usefully are limited by the coherence length and spatial coherence of the laser beam respectively.

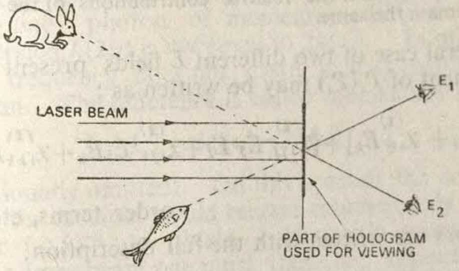


Fig. 21·14.

Viewing with a hologram carrying records of several objects

Colour holography has been developed ; as also three-dimensional holograms and holography of moving objects. But we shall not go into them. Interestingly, it has been stated that the use of holography will allow the storage of information in a three dimensional medium with such compactness that a 1 cm^3 crystal may accommodate the information contained in a library of a million books.

21·11. Nonlinear Optics. An electric field E would create in a molecule an electric polarization P according to the general relation :

$$P=\chi^{(1)}E+\chi^{(2)}E^2+\chi^{(3)}E^3+\ldots\ldots \qquad \ldots(21\cdot12)$$

where $\chi^{(1)}$, $\chi^{(2)}$, $\chi^{(3)}$, are polarizability tensors, which means that the contribution to P for each term could be in a direction other than E also. At the E magnitudes we have in usual light, only the first term has significant contribution. In that case the ratio P/E remains independent of intensity E, and we have what is called *linear optics*. In other words, if we have simultaneous fields E_1, E_2, etc. present, then the total result is the linear sum of the result due to individual fields. The 'principle of superposition' (see Ch. 11) is a recognition of this.

In the case of lasers we have such high energy densities that the corresponding E fields are able to produce substantial contributions to P from the higher order terms in Eq. 21·12. In such a situation P/E is not independent of intensity ; we have *nonlinear optics*.

Example 11·5. With certain units for P and E, let the three χ coefficients in Eq. 21·12 be of the order of 10^{-4}, 10^{-9} and 10^{-16} respectively. Evaluate the relative contributions of the three terms for (*a*) $E=1$ unit, (*b*) $E=10^4$ units, (*c*) $E=10^7$ units.

Solution. Simple substitution gives for the three cases :

(*a*) $1 : 10^{-5} : 10^{-12}$

(*b*) $1 : 10^{-1} : 10^{-4}$.

(*c*) $1 : 10^{+2} : 10^{+2}$.

Note : $\chi^{(1)}$, $\chi^{(2)}$, $\chi^{(3)}$, are dimentionally different. Hence their relative numerical values have no meaning by themselves ; they depend on the units for E and P. Thus, if the unit for E were taken 10^6—fold larger (numerical

E values 10^{-6} fold), then in the same example $\chi^{(1)}$, $\chi^{(2)}$, $x^{(3)}$ would be 10^{+2}, 10^{+3}, 10^{+5} respectively. But the relative contributions of the three terms of Eq. 21·12 will remain the same.

In a general case of two different E fields present (E_j and E_k) the ith component of P (P_i) may be written as :

$$P_i=[\chi_{ij}^{(1)} E_j+\chi_{ik}^{(1)} E_k]+[\chi_{ijj}^{(2)} E_j E_j+\chi_{ijk}^{(2)} E_j E_k+\chi_{ikk}^{(2)} E_k E_k]+\text{third order terms, etc.} \quad \text{...(21·13)}$$

where the tensors are written with the full description.

Nonlinear optics gives rise to many applications, out of which we shall outline only two here briefly.

Harmonic generation. Consider a laser beam of angular frequency ω. The electric field in it may be represented by $E=E_0 \cos \omega_0 t$. If this passes through a medium, whose response is represented by Eq. 21.12 (with the χ's as polarizability tensors of the medium), then because of high E_0 there will be a nonlinear response. Let us pay attention to the contribution of the *second term* alone and write it as

$$P^{(2)}=\chi^{(2)}E_0^2 \cos^2 \omega t$$

$$=\tfrac{1}{2}\chi^{(2)} E_0^2 +\tfrac{1}{2} \chi^{(2)} E_0 \cos 2\omega t \quad \text{...(21·14)}$$

The last result follows simply be expressing $\cos^2 \omega t$ as $\frac{1}{2}$ $(1+\cos 2\omega t)$. Thus, apart from a cos ωt dependent component of P, we get a static response (time independent), and one of angular frequency 2ω. In a similar way, the third term in Eq. 21·12 gives a contribution of angular frequency 3ω.

Again, consider two laser beams of angular frequencies ω_j and ω_k passing simultaneously through a medium. Then the second bracket in Eq. 21·13 gives

$$P_i^{(2)}=\chi_{ijj}^{(2)} E_{jo}^2 \cos \omega_j t+\chi_{ijk}^{(2)} E_{jo} \cos \omega_j t\, E_{ko} \cos \omega_k t+\chi_{ikk}^{(2)} E_{ko}^2+\cos \omega_k t \quad \text{...(21·15)}$$

In this the first and third terms give a static field plus two harmonics of frequencies $2\omega_j$ and $2\omega_k$. The middle term needs notice, it gives responses of frequencies $\omega_j+\omega_k$ and $\omega_j-\omega_k$.

One may think that occurrence of the harmonies will depend only on the medium having a significant $\chi^{(2)}E^2$ or $\chi^{(3)} E^3$ contribution. This is only partly true. A laser beam by itself does not produce harmonics; we need a medium with good $\chi^{(2)} E^2$ or $\chi^{(2)} E^3$ contribution. But another condition needs to be satisfied. That condition is *momentum matching*.

In producing the harmonic 2ω out of a laser beam ω the medium acts as an agency to absorb two photons of momentum $\hbar k(\omega)$ each and emitting one photon of momentum $\hbar k(2\omega)$. Now, due to colour dispersion $k(2\omega)$ is *not* equal to $2k(\omega)$. In other words two-fold angular frequency does not mean exactly two-fold momentum for the photon. The difference is called momentum mismatch

$$|\,\triangle k\,| = k(2\omega) - 2k(\omega) \qquad \text{...(21·16)}$$

($\hbar$ is conventionally omitted). On this account the second harmonic generation in the medium would remain coherent only over a length of the order $(\triangle k)^{-1}$; to obtain a harmonic generation which is directed like a laser beam one must therefore achieve $\triangle k \to 0$. This can be had in an anisotropic crystal. There this difference between ordinary and extraordinary components varies with direction (optical birefringence) and one may choose the direction such that this difference just balances the effect of colour dispersion.

With such an arrangement, about two thirds of input power at $\omega = 1.06\ \mu m$ was first converted to 2ω ($0.53\ \mu m$), and then an equal number of ω and 2ω photons were combined (in another crystal) to convert ever 80% of the laser power to the third harmonic power at 3ω ($0.353\ \mu m$).

The significance of momentum matching is that while algebra may indicate components of frequencies $0, \omega, 2\omega, 3\omega$, etc. being present simultaneously according to the coefficients, the physics of momentum conservation decides which of the components will dominantly show up in the experimental set up.

Example 21·6. Harmonic 2ω is to be generated out of a laser beam $\omega = 4\times10^{14}$ Hz. If the medium has refractive index 1.528 at ω and 1.542 at 2ω, deduce the momentum mismatch due to colour dispersion. Also estimate the lengh of the medium for which the 2ω beam will remain coherennt if the mismatch is not compensated.

Solution. We have

$$k = \frac{2\pi}{\lambda} = \frac{2\pi\nu}{c'} = \frac{\omega}{c}\, n$$

where c' is speed of light in the medium and n is the appropriate refractive index. Substitution leads to

$$\triangle k = \frac{8\times10^{14}\times1·542}{3\times10^{8}} - 2\,\frac{4\times10^{14}\times1·528}{3\times10^{8}}$$

$$\simeq 4\times10^{4}\ \text{m}^{-1} \qquad \odot$$

$$\Rightarrow \quad L \simeq (\triangle k)^{-1} \simeq 2\times10^{-5}\,\text{m}$$

Nonlinear Spectroscopy. It may be noted that $\chi^{(1)}$ for a medium as also for a molecule depends on frequency. For a medium this leads to dispersion, and for a molecule it leads to resonance. Our attention at present may be on $\chi^{(2)}$ or $\chi^{(3)}$. In the case of two beams of frequencies ω_L and ω_S we find that $\chi^{(2)}$ and $\chi^{(3)}$ are strongly dependent on the difference of ω_L and ω_S; they are maximum when the difference is equal to one of the eigenfrequencies of the molecule, which we may call ω_R. This is the basis of nonlinear spectroscopy for the determination of characteristic eigenfrequencies ω_R for sample molecules.

In conventional Raman effect we illuminate a sample at frequency ω_0, and find that the scattered beam has light of frequencies $\omega_0+\omega_R$ and $\omega_0-\omega_R$ besides ω_0. The light of shifted frequencies is very weak in general and it goes in all directions (is incoherent). But that is what is used in conventional Raman effect to learn about characteristic ω_R values. The advent of lasers has made very strong exciting sources, so that the Raman study becomes much faster. That is about conventional Raman study with lasers. We are not here talking about that. Nonlinear spectroscopy involves the study of coherent beams only.

Let ω_L refer to a strong laser beam, called the pump beam; and let ω_S refer to another laser beam which is of variable frequency (tunable) and is called the probe beam. Let both of them pass through the sample at appropriate mutual inclination (we need it for momentum conservation). If we now pay attention to the interaction represented symbolically by

$$P^{(3)}(\omega)=\chi^{(3)}_{SLL}\ E(w_S)\ |\ E(\omega_L)\ |^2 \qquad \ldots(21.17)$$

then we have what is called 4-wave mixing, which means ω_S, ω_L, and ω_L photons act together to produce an w photon. This ω could be $2\omega_L+\omega_S$ or $2\omega_L-w_S$. Also we may be making the measurements on either of these or on the ω_L or ω_S beams themselves as the frequency ω_S is varied. Sharp increase in the intensity of ω beam or sharp fall in the intensity of ω_L or of ω_S beam is observed at specific ω_S values. We relate this fact to the difference $|\ \omega_L-\omega_S\ |$ becoming equal to ω_R, and thus get the characteristic Raman frequencies.

Notice that the ω_L and ω_S beams are coherent; also that by suitable choice of relative inclinations the ω beam will also be coherent. In that sense laser Raman spectroscopy is not the same as conventional Raman spectroscopy using laser as a strong source.

The special significance is that the coefficient $\chi^{(3)}_{SLL}$ involves much more detail than just the frequency difference; it also depends on orientation of the molecule relative to the beams. It is therefore able to lead us to more details about the structure and symmetry of the molecule.

PROBLEMS

21·1. Distinguish between :

(*i*) 'Coherence between two sources' and 'Coherence of a given source'.

(*ii*) 'Temporal coherence' and 'spatial coherence'.

21·2. For a source radiating at $\lambda=5400A$ the coherence time τ is $\sim 2\times 10^{-10}$ sec. Deduce the order of magnitude values of (*i*) the coherence length L, (*ii*) the spectral half-width $\Delta\lambda$ of the radiation, (*iii*) the purity of the spectral line $(\lambda/\Delta\lambda)$, and (*iv*) Q of the source.

21·3. Explain the essential requirements for producing laser action. Outline how these requirements are usually obtained.

21·4. Define Einstein's coefficients A_{12} and B_{12}. Laser action is sometimes called inverted absorption. Explain. In what situations may A_{12}/B_{12} be small enough for laser action ?

21·5. If the energy gap between two levels is equivalent to radiation of $\lambda=5000\ A°$, the Boltzmann population ratio N_2/N_1 at room temperature comes to $\sim e^{-100}$. Check it. Compute the ratio for gap corresponding to $\lambda=1$ cm. Explain why masers were achieved earlier than lasers.

21·6. With reference to Fig. 21·4 (*b*) show that in the equilibrium state

$$\frac{dN_1}{dt}=-N_1(u_{13}B_{13}+u_{12}B_{12})+N_3(u_{13}B_{13}+\mathrm{A}_{13})+N_2(u_{12}B_{12}+\mathrm{A}_{12})=0$$

Now explain the following : (*a*) N_3 can never be $>N_1$, (*b*) to get $N_2>N_1$ the level 2 must be a metastable one, and (*c*) to make the stimulated emission dominant one must use a cavity resonator.

21·7. In a ruby laser (§ 21·8) the absorption band has width 800 $A°$ and the total emission of *Xe* source may be taken as spread over $\sim 3000A°$. Further, a transition $3\rightarrow 1$ is ~ 300 times more probable than $3\rightarrow 2$, and out of the re-emitted radiation of frequency ν_{12} roughly 10% may be re-absorbed by the Cr^{+3} ions. Estimate the percentage efficiency of the laser ($\lambda_{13}=5600A°$, $\lambda_{12}=6943A°$).

21·8. Compare glass lasers and gas lasers in their performance. What lasers are useful for wide band amplification ?

21·9. Explain how the spectral half-width of a radiation decreases in a laser to much smaller than its intrinsic value.

21·10. Describe how spatial coherence leads to high directionality of a laser beam. Light from a 2·5 *mw* laser source of aperture 1·8 cm dia and $\lambda=5000\ A$ is focussed by a lens of $f=20$ cm. Deduce the area and intensity of the image.

21·11. State the principle of holography. Why is a laser beam needed for it ? Can one plate record more than one holograms ?

21·12. In a laser Raman experiment

$$\omega_L = 4{\cdot}410 \times 10^{16} Hz, \quad \omega_S = 4{\cdot}291 \times 10^{16} Hz$$

and the refractive indices of the medium are

$$n(\omega_L) = 1{\cdot}4129, \; n(\omega_S) = 1{\cdot}4100, \; n(2\omega_L - \omega_S) = 1{\cdot}4198$$

Deduce the angle between the ω_L and ω_S beams for the best response, and thc direction of the $2\omega_L - \omega_S$ beam relative to the ω_L beam, using conservation of momentum.

22

Crystal Structure and Diffraction by Crystals

Almost all solids have regularity of atomic arrangements, *i.e.*, they are crystalline. (The few exceptions are called 'glasses', or glassy solids). In some cases this regularity extends upto only a few thousand atoms on each side, so that the regularity of arrangements is not seen in the external looks. In other cases, like cane sugar or common salt, the regularity of atomic arrangement extends to $\sim 10^7$ or more atoms in each direction very frequently, so that individual crystals have dimensions of a few mm on each edge. Then the regularity of flat faces with definite angles becomes obvious.

Study of the crystal structure of different solids is itself a very important aspect of solid state chemistry and physics. However, presently our interest is to see how crystalline structure of solids is used as a diffraction grating. The closest *ruled* gratings have a grating space $e \approx 10^{-4}$ cm$=10000\ A$, while X-rays have wavelengths of the order of a few angstroms. Bragg discovered that the regularity of atomic arrangements in crystals had a spacing of just the right order for determination of wavelengths of X-rays. We shall later see that even beams of electrons, neutrons, etc. show wave-like behaviour and diffraction from crystals can be used to measure the wavelengths which they show up. For a clearer understanding we shall include some details of crystal structure before we proceed to describe crystal spectrography.

22·1. Space Lattices. A lattice is a framework of geometrical points in three dimensional space described by the *translation operation* :

$$T=n_1\mathbf{a}+n_2\mathbf{b}+n_3\mathbf{c} \qquad \text{...(22·1)}$$

where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are three *unit vectors*, and n_1, n_2, n_3 take all possible integral values. The parallelopiped formed by $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ as the concurrent edges is called the *unit cell* of the lattice. A unit cell repeated with

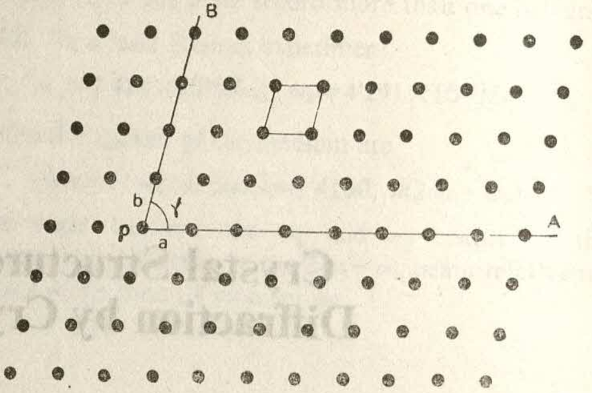


Fig. 22·1.
A two-dimensional space lattice

translations T fills all space. Fig. 22·1 shows a two-dimensional space lattice, and Fig. 22·2 shows a three-dimensional unit cell.

Seven Systems of Crystals.

On the basis of the shape of the unit cell all crystals are classified into 7 *systems*. Their names and characteristics are given in Table 22·1. The least symmetry is in triclinic crystals, and the largest symmetry is in cubic crystals. We shall limit our discussion largely to the *cubic system* alone. This is not only the simplest, but is also amongst the commonest in nature. Amongst the elements, well over half crystallise in the cubic system.

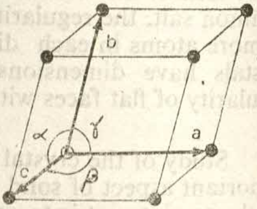


Fig. 22·2.
A Unit cell in 3 dimensional lattice

TABLE 22·1
The Seven Systems of Crystals

Name of the System	*Relation of lengths of Axes*	*Relation of Angles between Axes*
Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$
Monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90° \neq \beta$
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$
Hexagonal	$a = b \neq c$	$\alpha = \beta = 90°, \gamma = 120°$
Trigonal	$a = b = c$	$\alpha = \beta = \gamma \neq 90°$
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90°$

It may be noted that with $a \neq b \neq c$, if two angles become equal, they have to be 90°, otherwise no three dimensional repetitive arrays can be formed. Similarly with two axes equal $(a=b)$, the only repetitive structures possible are when two angles are 90° and the third one is either 90° or 120°. Finally with $a=b=c$ all three angles also have to be equal, otherwise 3-dimensional repetitive structure cannot be formed. (We omit proofs for these). It is for these reasons that the geometrical variety of unit cells is limited to just seven.

Bravais Space Lattices. Consider the three unit cells shown in fig. 22.3. All three have a cubic shape. The first has lattice points at the corners only; the second has an extra lattice point at the centre of *each* face ; the third has an extra point at the centre of its volume. These fall under *one system* (cubic system), but *three different space lattices.*

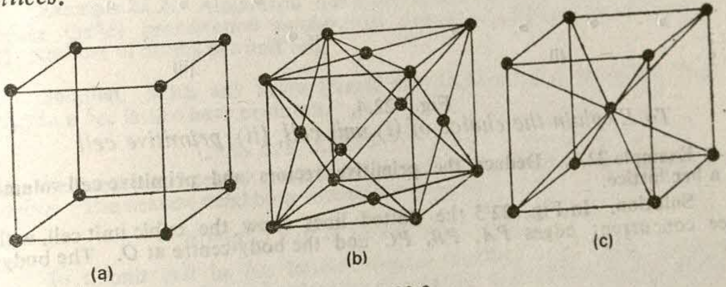


Fig. 22.3.

Unit Cells in simple cubic (Cubic P), face-centred cubic (cubic F) and body-centred cubic (cubic I) space lattices of Bravais

Bravais showed that there could be in all 14 different lattices under the seven systems. When there is no extra lattice point in the unit cell besides the corners, the cell is called a *primitive cell* (*P*). When there are lattice points at the face-centres or at the body centre, symbols *F* and *I* respectively are used. Symmetry considerations decide which of these can be had in each system. For instance, in the cubic system we have *P*, *F* and *I* lattices. In triclinic system only primitive cells (*P*) can occur.

Cubic *P*, cubic *F*, and cubic *I* lattices are also referred to as sc (simple cubic), fcc (face-centred cubic) and bcc (body-centred cubic) lattices respectively. One way of looking at the bcc lattice is to consider it as one sc lattice plus *one* more shifted by $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Similarly fcc lattice is equal to one sc lattice plus *three* more shifted by $(0, \frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 0, \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2}, 0)$. [All coordinates here are expressed in units of *a*].

Unit Cell and Primitive Cell. In Fig. 22.4(*i*) we have a 2-dimensional lattice. We could choose *ABCD*, or *ABEC* or *ABFE*.. as the unit cell, with appropriate translation (Eq. 21.1) to fill all space. The choice ? That which represents the *maximum symmetry* available is chosen as the unit cell. *ABCD* is thus the choice.

In Fig. 22.4 (*ii*) is another situation. The *smallest* unit is *ABEF*, but *smallest with maximum symmetry is ABCD*. In this case *ABCD* is chosen as the *unit cell*. *ABEF* is called the *primitive cell*. A unit cell may have multiple lattice points per cell ; a primitive cell has just one.

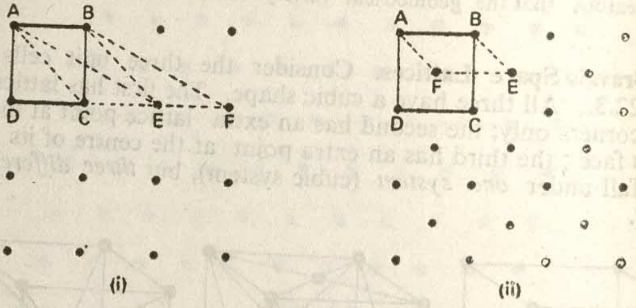


Fig. 22.4.
To Explain the choice of (*i*) *unit cell,* (*ii*) *primitive cell*

Example 22.1. Deduce the primitive vectors and primitive cell volume for a bcc lattice.

Solution. In Fig. 22·5 the dotted lines show the cubic unit cell, with three concurrent edges *PA*, *PB*, *PC* and the body-centre at *O*. The body-

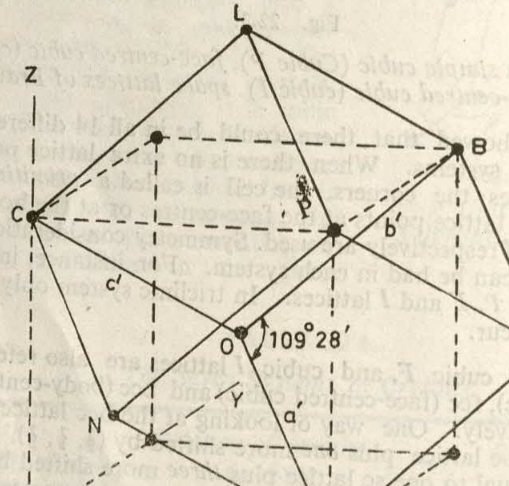


Fig. 22·5.
Primitive cell for bcc lattice

centres of neighbouring unit cells meeting at *P* are *L*, *M* and *N*. The vectors *OA*, *OB*, *OC* are the primitive cell vectors ; you could as well take *PL*, *PM*, *PN* as the primitive cell vectors.

The coordinates required are

$$O\left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right) \quad A\,(1, 0, 0) \quad B\,(1, 1, 1), \quad C\,(0, 0, 1)$$

So the vectors are (in units of $\mathbf{a}$)

$$OA \equiv \mathbf{a}' = \left(\tfrac{1}{2}i - \tfrac{1}{2}j - \tfrac{1}{2}k\right), \quad OB \equiv \mathbf{b}' = \left(\tfrac{1}{2}i + \tfrac{1}{2}j + \tfrac{1}{2}k\right), \quad OC \equiv \mathbf{c}' = \left(\tfrac{1}{2}i - \tfrac{1}{2}j + \tfrac{1}{2}k\right)$$

These lead to

$$a' = \frac{\sqrt{3}}{2}a.$$

$$\cos\gamma = \frac{\mathbf{a}'\cdot\mathbf{b}'}{a'b'} = \frac{\frac{1}{4} - \frac{1}{4} - \frac{1}{4}}{3/4} = -\tfrac{1}{3} \Rightarrow \gamma = 109^\circ\ 28'$$

$$\text{volume} = \mathbf{c}'.\mathbf{a}' \times \mathbf{b}' = \tfrac{1}{2}a^3 \text{ (check it).}$$

Example 22.2. Alpha iron has a *bcc* lattice with lattice constant 2.86A°. Deduce the (*i*) *coordination number*, (*ii*) nearest neighbour distance, and (*iii*) (*iii*) Number of atoms per unit cell.

Solution. With any atom treated as reference atom, the nearest neighbours in a *bcc* lattice have positions

$$\left(\pm\tfrac{1}{2}, \pm\tfrac{1}{2}, \pm\tfrac{1}{2}\right) a$$

Thus the number of nearest neighbours is $2 \times 2 \times 2 = 8$. This is the *coordination number*. The nearest neighbour distance is given by

$$d^2 = (a/2)^2 + (a/2)^2 + (a/2)^2 = a\sqrt{3}\,/2$$

In a unit cell in *bcc* lattice there are 8 atoms at the corner +1 at the body-centre. Since each corner atoms is shared by 8 unit cells, there are $(8/8) + 1 = 2$ atoms per unit cell.

22.2. Crystal Structure. The lattice is only an array of *geometrical* points or a framework in which atoms are to be arranged. In an actual crystal, each point has arround it an assembly of atoms, which is called *the basis*. In Fig. 22.6 at (*a*) we show *the lattice*, at (*b*) we show *the basis* comprising two different atoms, and at (*c*) we show the *crystal structure* obtained by introducing the basis at each lattice point. Thus one may symbolically write

$$\text{Lattice} + \text{basis} = \text{crystal structure}$$

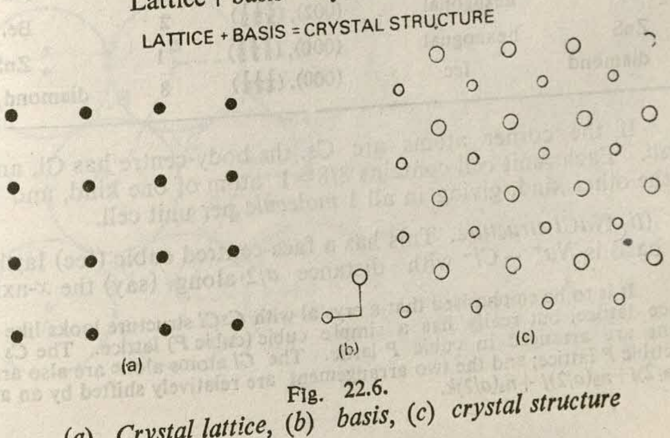


Fig. 22.6.

(*a*) *Crystal lattice*, (*b*) *basis*, (*c*) *crystal structure*

In other words, the basis repeated according to a translation operation (Eq. 22.1) gives us the crystal structure. Note that for a compound containing several atoms in each molecule (and also water of crystallization) the crystal *structure* will in general be quite complicated, but if attention is centred on equivalent atoms of one kind only, we get in each case the Bravais *lattice*. There are only 14 such *lattices*, though the number of crystal *structures* is unlimited.

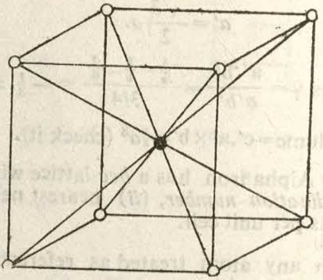


Fig. 22.7.
The CsCl structure; one unit cell is shown

Some important crystal structures are listed in Table 22.2, with typical examples. Description of these structures follows.

(*i*) *CsCl Structure*. The unit cell for this is shown in Fig. 22.7.

TABLE 22.7

Some Typical Crystal Structures

Structure	*lattice*	*basis*	*molecules per unit cell*	*Examples*
CsCl	sc	(000), $(\frac{1}{2}\frac{1}{2}\frac{1}{2})$	1	CsCl
NaCl	fcc	(000), $(\frac{1}{2}00)$	4	NaCl, KCl
hcp	hexagonal	(002), $(\frac{2}{3}\frac{1}{3}\frac{1}{2})$	2	Be, Mg
ZnS	hexaognal	(000), $(\frac{2}{3}\frac{1}{3}\frac{1}{2})$	1	ZnS
diamond	fcc	(000), $(\frac{1}{4}\frac{1}{4}\frac{1}{4})$	8	diamond, Ge, Si

If the corner atoms are Cs, the body-centre has Cl, and *vice-versa*. Each unit cell contains $8/8=1$ atom of one kind, and 1 atom of the other kind, giving in all 1 *molecule* per unit cell.

(*ii*) *NaCl structure*. This has a face-centred cubic (fcc) lattice and the basis is $Na^+ - Cl^-$ with distance $a/2$ along (say) the x-axis. In

It is to be emphasised that a crystal with *CsCl* structure looks like having a bcc lattice, but really has a simple cubic (cubic *P*) lattice. The *Cs* atoms alone are arranged in cubic *P* lattic. The *Cl* atoms alone are also arranged in cubic *P* lattice; and the two arrangement are relatively shifted by an amount $n_1(a/2)i+n_2(a/2)j+n_3(a/2)k$.

Fig. 23.8 if solid dots show Na^+ ions, the hollow circles show Cl^-, or vice-versa.

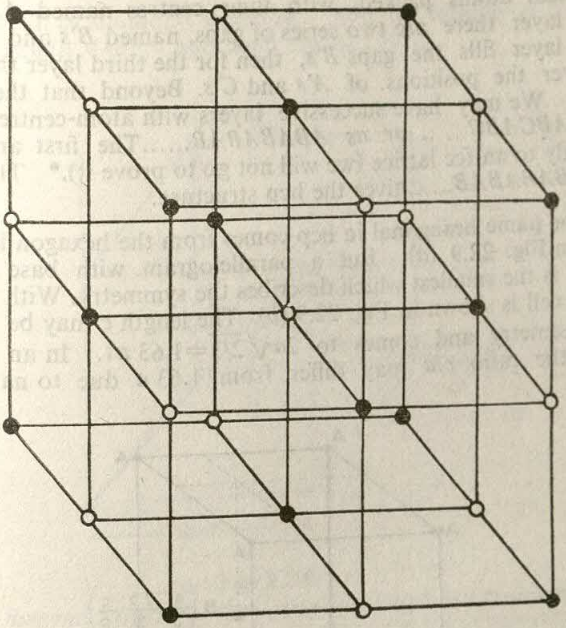


Fig. 22.8.

The NaCl structure. One unit cell is shown.

If you ignore the Na^+ and Cl^- difference the $NaCl$ lattice looks like a simple cubic one. But one kind of atoms alone show the fcc lattice. If lattice parameter is a, the nearest neighbour distance to $a\sqrt{2}$ for atoms of the same kind and $a/2$ for atoms of the opposite kind. The coordination number refers to nearst neighbours (here of the other kind) and is 6.

(*iii*) *hcp* (*hexagonal close-packed*) *structure.*

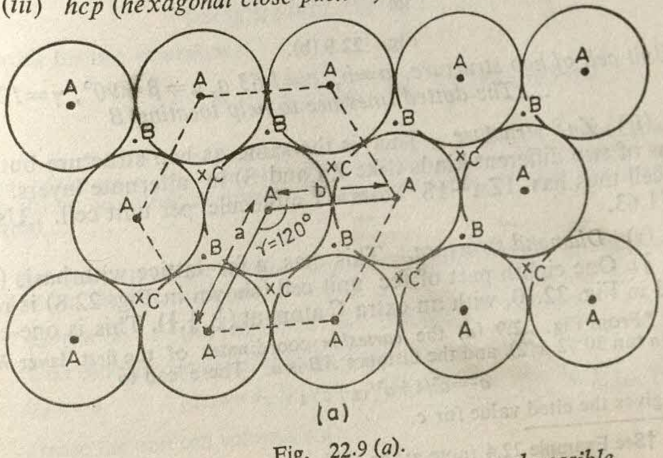


Fig. 22.9 (*a*).

One layer of atoms (*A*) *in hcp structure, and possible locations of atoms in the next layer* ; *B or C*

This may be understood from Fig. 22.9 (*a*), showing one layer of spherical atoms packed, with their centres named *A*. For the second layer there are two series of gaps, named *B's* and *C's*. If the second layer fills the gaps *B's*, then for the third layer there will be gaps over the positions of *A's* and *C's*. Beyond that there are two choices. We may have successive layers with atom-centres arranged as *ABCABCABC*...... *or as ABABABAB*......The first arrangement leads only to an fcc lattice (we will not go to prove it).* The arrangement *ABABABAB*......gives the hcp structure.

The name hexagonal in hcp comes from the hexagon base shown dotted in Fig. 22.9 (*a*). But a parallelogram with base $a=b$ and $\gamma=12$ ° is the smallest which describes the symmetry. With $\alpha=\beta=90°$ the unit cell is shown in Fig. 22.9 (*b*). The length *c* may be calculated from geometry and comes to $2a\sqrt{2/3}=1.63\ a$†. In an actual hcp crystal the ratio c/a may differ from 1.63 *a* due to nature of the bonds.

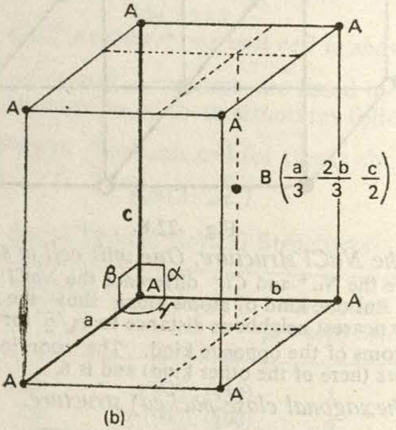


Fig. 22.9 (b).

Unit cell of hcp structure. $a=b$, $c=1.63\ a$, $\alpha=\beta=90°$, $\gamma=120°$ .....*The dotted lines are to help locating B*

(*iv*) *ZnS structure.* This is the same as hcp structure but with atoms of two different kinds (like Zn and S) in alternate layers. The unit cell thus has 1Zn+1S atom=1 molecule per unit cell. Usually $c/a \neq 1.63$.

(*v*) *Diamond structure.* This has a fcc lattice with basis (000), $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. One eighth part of fcc unit cell (shown in Fig. 22.8) is reproduced in Fig. 22.10, with an extra *C* atom at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. This is one-eighth

*From Fig. 22.9 (*a*) the *cartesian* coordinates of the first layer *B* are $(a/2,\ a \tan 30°/2,\ c/2)$, and the distance *AB* is *a*. These lead to

$$a^2=a^2/4+a^2/12+c^2/4$$

That gives the cited value for *c*.

---

†See Example 22.4 (note at the end).

fothe unit cell of diamond structure. Note the tetragonal *C-C* bonds with bond length $a\sqrt{3}/4$ and bond angles 109°28.

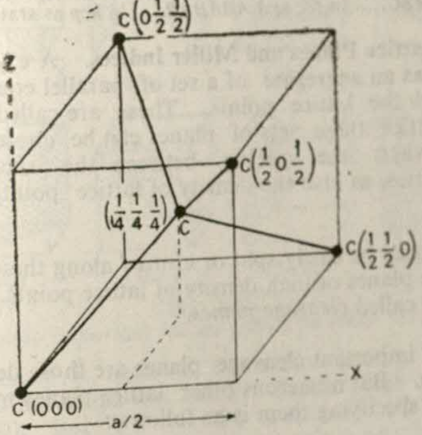

$c(0\frac{1}{2}\frac{1}{2})$
z
$c(\frac{1}{2}\,0\,\frac{1}{2})$
$(\frac{1}{4}\frac{1}{4}\frac{1}{4})$ C
$c(\frac{1}{2}\frac{1}{2}\,0)$
x
C (000)
a/2

Fig. 22·10.

Diamond Structure. One-eighth of a unit cell is shown.

Example 22·3. The density of sodium chloride is 2·18 g/cc. Given that the crystal has fcc space lattice deduce the lattice constant *a*. (Avogadro number$=6{\cdot}02\times10^{23}$ per g-mole).

Solution. Volume of unit cell is a^3 cc and mass in each unit cell is therefore $a^3\times2{\cdot}18$ g.

Now molecular weight of *NaCl* is 58·5, and in fcc lattice there are 4 molecules per unit cell. Hence mass in each unit cell is

$$(4\times58{\cdot}5)\div(6{\cdot}02\times10^{23})\text{g}$$

Equating the two expressions.

$$a^3\times2{\cdot}18=\frac{4\times58{\cdot}5}{6{\cdot}02\times10^{23}} \Rightarrow a=5{\cdot}63\times10^{-8}\text{ cm}$$

Example 22·4. Packing fraction in crystals is defined as the ratio of volume of the atoms in a unit cell and the volume of unit cell. Compute it for (i) sc, (ii) fcc, and (iii) hcp structure of elemental solids, treating the atoms as spherical.

Solution. If atomic diameter is *d*, the volumes occupied by the atoms in a unit cell in the three cases are :

(*i*) $[\pi d^3/6]\times1$ (*iii*) $[\pi d^3/6]\times4$ (*iii*) $[\pi d^3/6]\times2$

Unit cell parameters for the cass are :

(*i*) $a=d$, (*ii*) $a=d\sqrt{2}$, (*iii*) $a=b=d$, $c=2d\sqrt{2/3}$, $\gamma=120^\circ$

Hence the unit cell volumes are :

(*i*) d^3 (*ii*) $2\sqrt{2}\,d^3$, (*iii*) $\sqrt{2}\,d^3$

The required packing fractions therefore are

(*i*) $\pi/6$ (*ii*) $\pi/3\sqrt{2}$ (*ii*) $\pi/3\sqrt{2}$

[Note that *fcc* has the same packing fraction as *hcp*. Only the arrangements are *ABCABC*......in *fcc* and *ABABAB*......in *hcp* as stated before]

22·3. Lattice Planes and Miller Indices. A crystal lattice may be considered as an aggregate of a set of parallel equi-distant planes passing through the lattice points. These are called *lattice planes.* For a given lattice these sets of planes can be chosen in a number of different ways; the spacing between the successive planes accordingly varies, as also the density of lattice points per unit area in each plane.

A crystal can be easily split or *cleaved* along these lattice planes, particularly the planes of high density of lattice points. Hence lattice planes are also called *cleavage planes.*

The most important cleavage planes are those defined by (**a**, **b**), (**b**, **c**) and (**c**, **a**). But numerous other lattice planes may be defined. The method of specifying them is as follows :

(*i*) Let the intercepts by the given plane on the three crystal axes be in the ratio $pa : qb : rc$, where a, b, c are the corresponding unit vector lengths, and p, q, r are integers.

(*ii*) Take the reciprocals of p, q, r.

(*iii*) Get the smallest possible integers h, k, l such that

$$h : k : l = p^{-1} : q^{-1} : r^{-1}. \qquad ...(22{\cdot}2)$$

Now these numbers h, k, l are known as *Miller indices* of the given set of planes, and we specify the plane as $(h\ k\ l)$. In terms of coordinate geometry the family of $(h\ k\ l)$ planes is described by

$$\frac{hx}{a}+\frac{ky}{b}+\frac{lz}{c}=n \quad (n=\text{integer}) \qquad ...(22{\cdot}3)$$

For planes parallel to the xy plane we have $p=q=\infty$ and r finite. Therefore, the Miller indices are (001). Similarly, the indices (0 1 0) and (1 0 0) represent lattice planes parallel to zx and yz planes respectively.

Consider the (1 1 0) plane. It is parallel to z-axis and intercept along x and y axes are equal in terms of the respective primitives. Similarly, for (2 1 0) plane the intercept along x-axis is 1/2 times that along the y-axis. For (3 2 4) plane the x-,y-,z-intercepts are in the proportion 1/3, 1/2, 1/4, which means 4 : 6 : 3.

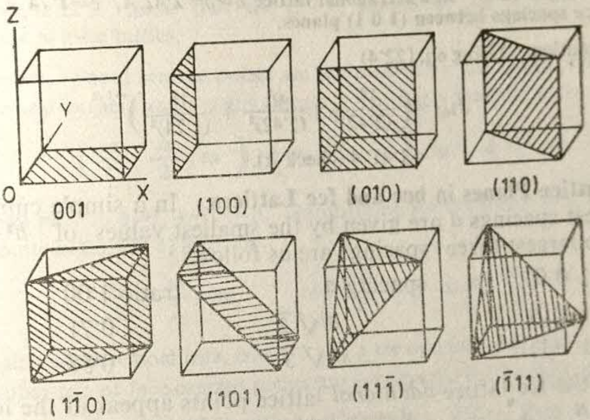


Fig. 22·11.
Some important lattice planes in cubic crystals

If an intercept is negative the corresponding Miller index is written with a bar at the top, like $\bar{3}$. Thus a plane (1 $\bar{3}$ 0) is not the same as (1 3 0). But ($\bar{1}$ $\bar{3}$ 0) is the same as (1 3 0), because *if all Miller indices are reversed in sign we get only a parallel plane*, as may be seen from Eq. (22·3).

Example 22·5. In a triclinic crystal* a cleavage plane makes intercepts 2·93, 4·47 and 2·35 units along the three crystallographic axes, the corresponding primitives being 3·05, 6·99 and 4·90 *A*. Deduce the Miller indices of the cleavage plane.

Solution. $p : q : r = \frac{2 \cdot 93}{3 \cdot 05} : \frac{4 \cdot 47}{6 \cdot 99} : \frac{2 \cdot 35}{4 \cdot 90}$

$= \cdot 961, \cdot 640 : \cdot 480 = 6 : 4 : 3$

$\therefore \quad p^{-1} : q^{-1} ; r^{-1} = \frac{1}{6} : \frac{1}{4} : \frac{1}{3} = 2 : 3 : 4$

Hence the Miller indices of the plane are (2 3 4).

Spacing between Lattice Planes. If we take successive planes with Miller indices* $\{h\ k\ l\}$, then for primitive lattices in which $\alpha = \beta = \gamma = 90^\circ$ the spacing between the planes is given by

$$d_{hkl} = \left[\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2} \right]^{-1/2} \qquad \text{...(22·4)}$$

We shall not prove the formula here. In the case of the cubic system, $a = b = c$. Hence for cubic *P* lattice eq. (22·4) simplifies to

$$d_{hkl} = \frac{a}{(h^2 + k^2 + l^2)^{1/2}} \qquad \text{...(22·5)}$$

---

* The specifications tricinic or orthorhombic are unimportant in these examples. But when the spacing between the planes is to be considered, these become meaningful.

Example 22·6. In a tetragonal lattice $a=b=2{\cdot}42\ A$, $c=1{\cdot}74\ A$. Deduce the lattice spacings between (1 0 1) planes.

Solution. Using eq. (22·4)

$$d_{101}=\left(\frac{1^2}{(2{\cdot}42)^2}+\frac{0^2}{(2{\cdot}42)^2}+\frac{1^2}{(1{\cdot}74)^2}\right)^{-1/2}$$
$$=1{\cdot}41\ A \text{ (check it).}$$

Lattice Planes in bcc and fcc Lattices. In a simple cubic lattice the largest spacings d are given by the smallest values of $h^2+k^2+l^2$. Thus the largest three spacings are as follows :

planes {1 0 0}	spacing a	ratio 1·00
{1 1 0}	$a/\sqrt{2}$	0·71
{1 1 1}	$a/\sqrt{3}$	0·58

In the fcc lattice* *additional* lattice points appear at the locations $\left\{\frac{a}{2}, \frac{a}{2}, a\right\}$. Now the {1 0 0} and {1 1 0} planes do not cover the face-centred points and we have to draw additional planes half-way through in each case. However, the (1 1 1) planes take care of the additional points without introducing extra planes. A look at Fig. 22·11, will make this clear Mathematically the decision follows from Eq. (22·3), which for cubic lattic becomes

$$hx+ky+lz=na \qquad \ldots(22.6)$$

The {a/2, a/2, 0} points do not satisfy this for the {1 0 0} and {1 1 0) planes; they do for the {1 1 1} planes.

Some authors specify the half-way planes by using Miller indices {2 0 0} and {2 2 0}, which is a modification of step (*iii*) to define h, k, l. This conveniently gives the spacings of Miller planes in fcc lattice by use of eq. (22·5). The lowest three $h^2+k^2+l^2$ now correspond to the planes as listed below :

planes {1 1 1}	spacing $a/\sqrt{3}$	ratio 1·00
{2 0 0}	$a/2$	0·87
{2 2 0}	$a/2\sqrt{2}$	0·61

For the *bcc* lattice, one may see similarly that additional planes arise half-way in the cases of {1 0 0} and {1 1 1} orientations, but not in the case of {1 1 0}. Hence the first two planes become {2 0 0} and {2 2 2} and hence planes {1 1 0} have the largest spacing. The three largest spacings are as follows :

plane {1 1 0}	spacing $a/\sqrt{2}$	ratio 1·00
{2 0 0}	$a/2$	0·71
{2 2 2}	$a/2\sqrt{3}$	0·41

---

* **The curly brackets{ } indicate all equivalent planes *e.g.*, {1 1 2} includes the four planes (1 1 2) ; ($\bar{1}$ 1 2) ; (1 $\bar{1}$ 2) ; (1 1 $\bar{2}$). These are differently oriented, but for all of them the spacing d_{hkl} is the same.**

Example 22·7. Show that the (1 1 0) planes do not cover all the face centre points in a fcc lattice.

Solution. Faced centred points are represented by $(ra/2, sa/2, t/a)$ where r and s are odd integers and t is any integer. The curly brackets mean

$$\left(\frac{ra}{2}, \frac{sa}{2}, ta\right), \left(ta, \frac{ra}{2}, \frac{sa}{2}\right), \text{ and } \left(\frac{sa}{2}, ta, \frac{ra}{2}\right)$$

(110) planes refer to $h=1$, $k=1$ $l=0$. Substitution for the three sets of face-centred points in Eq. (22·6) leads to

$$\frac{ra}{2}+\frac{sa}{2}=na, \qquad ta+\frac{ra}{2}=na, \text{ and } \frac{sa}{2}+ta=na$$

Only the first of these hold true, (since r and s are odd integers) which means that the other sets of face-centred points are *not* covered by the (110) planes.

22·4. Diffraction of X-rays by Crystals. When X-rays fall on a crystal, each electron of each atom scatters a part of the beam, and we have to consider the sum-total of the scattered radiation. This may be seen in three steps :

(*i*) The amplitude of the scattered wave resulting from Z different electrons *of a given atom* is not just Z-fold the amplitude of scattered wave per electron; it is AZ, where A is called the *atomic structure factor*. It depends on the distribution of electrons in the atom, λ of the X-rays used and the direction of observation relative to the incident light.

(*ii*) The amplitude of the scattered wave resulting from the different atoms in a unit cell of the crystal is not just $\Sigma A_i Z_i$ summed over the atoms in the unit cell, but is $f\Sigma A_i Z_i$, where f is called the *crystal structure factor*. It depends on the relative placements of the atoms in a unit cell, on λ of the X-rays used, and on the direction of observation.

(*iii*) The amplitude of the scattered wave from a crystal as a whole is then the resultant arising from the *lattice arrangement* of different unit cells. For some chosen directions the contributions from all unit cells add-up in phase. These form the maxima, and their positions depend on the lattice parameters, the λ value, and direction of observation.

Laue followed this procedure. For a beam passing through a single crystal, the transmitted light shows maxima in the form of **Laue spots** formed on a screen. From their *positions* and *intensities* one deduces not only a, b, c, but also structure details within the unit cell, and yet further the details of electron charge distribution in the atoms.

But we will use a short-cut, used by **Bragg** which centres on an examination of the lattice.

Bragg's Law. In Fig. 22·12 we show a set of parallel atomic planes*, an incident beam at angle θ from the plane, and a direction of observation θ'. If θ' corresponds to specular reflection at the atomic planes [*i.e.*, $\theta'=\theta$, and the incident ray, scattered ray, and normal to the lattice planes are in one plane], then it is simple to see that the scattered waves *from all the atoms in a given plane* reach the observer in the same phase, and therefore reinforce.

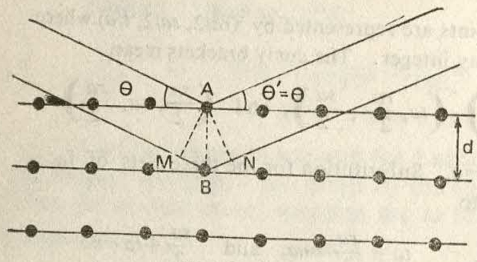


Fig. 22·12.
Bragg's Law

Next, the scattered waves from *successive planes also must reinforce.* The relation for this is deduced from Fig. 22·12. The incident wavefront is AM, and in the scattered beams from successive planes the path is common beyond AN. Hence the path difference between waves reflected from successive planes is $BM+BN=2d \sin \theta$. For reinforcement, this must be an integral multiple of λ. Hence

$$\boxed{2d \sin \theta = n\lambda} \qquad \text{Bragg's Law} \qquad \ldots(22.7)$$

Thus specular reflection condition *plus* Bragg's law means that waves *from all unit cells in all the lattice planes* reach the observer in agreement of phase.

An ordinary mirror behaves with visible light exactly the same way—each atom of the mirror being the centre of scattering. But for visible light the penetration is not large and $\lambda \gg d$. Hence Bragg's condition plays no role and light of all wavelengths shows a maximum at $\theta'=\theta$.

If a crystal belongs to a compound, then each 'dot' in Fig. 22·12. represents a unit cell, with atoms of different kinds arranged according to the crystal structure. One way to handle this is to first sum-up for a unit cell and then for the whole lattice. Another way is to first sum-up for the whole lattice for *one kind* of atoms, and then sum-up for atoms of different kinds. We shall use the second method, and consider only a diatomic crystal. Just two kinds of situations arise : (*i*) for some (h k l) values the planes may cover both the kinds of atoms, and (*ii*) for other (h k l) values the planes covering the other kind of atoms lie half-way between those causing the first kind of atoms.

As one example, in KCl (whose structure is of NaCl type, Fig. 22·8) the {1 1 1} planes alternately contain K^+ alone and Cl^- alone. Therefore for odd values of n the waves from K^+ planes and Cl^- planes have phase difference an *odd* multiple of π. Since K^+ has the same number of electrons as Cl^- the waves from K^+ planes and Cl^- planes totally cancel out for odd orders. In KBr the odd orders are weaker than even orders. But {1 0 0} planes for example, contain both K^+ and Cl^- (or Br^-) ions, and for these planes all the orders appear.

Bragg Spectrometer. The essential parts of a Bragg spectrometer are shown in fig. 22.13. *AB* is a narrow straight channel between two brass blocks. This serves as a narrow 'slit' (⊥ʳ to plane of paper) as well as 'collimator' for the incident *X*-rays. *C* is a single crystal, usually of NaCl with (1 0 0) face. *I* is an ionisation chamber*.

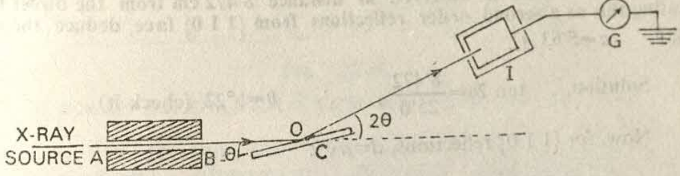


Fig. 22.13.

Schematic of a Bragg Spectrometer

The ionisation chamber and the crystal can be rotated about an axis at O, perpendicular to the plane of paper, in such a way that the angular relation θ and 2θ as shown in the diagram is always maintained.

As the angle 2θ is changed, the current through the galvanometer *G* shows some peaks, as represented schematically in fig. 22.14. This figure shows two close lines in first and second order. From the observed 2θ, and known d_{100}, along with suitable *n*, we can calculate λ values for the two lines. The lines are actually superposed over a continuous spectrum of *X*-rays. From this the intensity distribution in the continuous spectrum can also be deduced.

It will thus be seen that crystals provide natural (three-dimensional) gratings for the analysis of an *X*-ray spectrum. The maximum wavelength for which a crystal can be thus used is obviously limited to 2*d*.

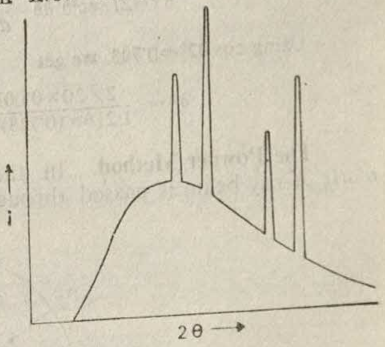


Fig. 22.14.

A typical Bragg spectrum

Example 22·8 Calculate the glancing angle on the cube face {1 0 0} of a rock-salt crystal (a=2·814 A) corresponding to second order reflection for *X*-rays of λ=0·710*A*.

Solution. For {1 0 0} planes, the spacing *d* is equal to *a*. Using eq. (22·7)

$$2\times 2{\cdot}814\times 10^{-8}\sin\theta = 2\times 0{\cdot}710\times 10^{-8}$$

$$\theta = \sin^{-1}\left[\frac{0\ 710}{2{\cdot}814}\right] = 14^\circ 22'.$$

---

* This may be replaced by a photographic film, in which case the record comes in one exposure and θ is determined by the geometry of the system (See example 22·9).

Example 22·9. In a Bragg spectrometer a photographic film is set perpendicular to the incident beam and at a distance 25·0 cm from the axis of the NaCl crystal. A line is observed at distance 8·472 cm from the direct beam. Treating this as a second order reflections from $\{1\ 1\ 0\}$ face, deduce the wavelength. $[a=5{\cdot}63\ A]$.

Solution. $\tan 2\theta=\dfrac{8{\cdot}472}{25{\cdot}0}$, $\quad\therefore\quad \theta=9°22'$ (check it)

Now, for $\{1\ 1\ 0\}$ reflections, $d=a\sqrt{2}$

$$\therefore\quad \lambda=\frac{2d\sin\theta}{n}=\frac{2\times 5\;63\times\sin 9°22'}{\sqrt{2}\times 2}=0{\cdot}651\ A$$

Example 22·10. In Bragg reflection, deviation of the beam is 2θ. Deduce the separation for beams of $\lambda=0{\cdot}810A$ and $0{\cdot}815A$ for second order reflection for planes of $d=1{\cdot}216A$, if a 20 cm camera is used.

Solution. Using eq. (22·7)

$$2\times 1{\cdot}216\times\sin\theta=2\times 0{\cdot}810 \quad\Rightarrow\quad \theta=42°\ \text{(check it)}$$

Also, differentiating eq. (22·7), we get

$$\delta\theta=\frac{n}{2d\cos\theta}\,\delta\lambda \qquad \text{...(22·8)}$$

Now, $x=f\tan 2\theta$. Hence

$$\delta x=2f\sec^2\theta\ \delta\theta=\frac{nf}{d\cos^3\theta}\,\delta\lambda$$

Using $\cos 42°=0{\cdot}743$, we get

$$\delta x=\frac{2\times 20\times 0{\cdot}005}{1{\cdot}216\times(0{\cdot}743)^3}=0{\cdot}43\ \text{cm}$$

⊙

The Powder Method. In this method (fig. 22.15) a *monochromatic X*-ray beam is passed through a pin hole (not slit) in a block of brass, and allowed to fall on a *powder* specimen. A photographic film-strip is put along a circle concentric with the specimen and the pattern

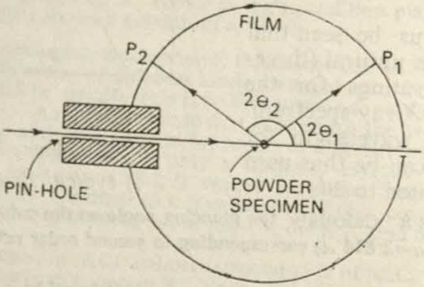


Fig. 22.15.

Powder Camera for analysis of crystal structure

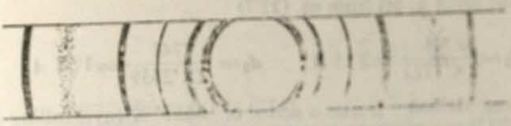


Fig. 22.16.

A powder pattern from X-rays on a cylindrical film strip

recorded is like the one shown typically in fig. 22.16. (The central spot is masked off). Alternatively, one could put a plane film on the front side alone, which gives a pattern of concentric circles, (like Fig. 22.20).

A 'powder' is an assembly of a large number of microcrystals oriented in all possible directions. Let d_1, d_2...be the prominent lattice spacings* for the given specimen. Then for the single wavelength of incident *X*-rays chosen for the experiment there will be a few angles θ_1, θ_2...... etc. for which Bragg equations (22.7) is satisfied with $n=1$ or 2. From the aggregate of micro-crystals only those crystals which have these proper orientation give Bragg reflections, which occur at the appropriate 2θ angles from the incident beam. Thus there are cones of Bragg reflection with semi-angles $2\theta_1$, $2\theta_2$, $2\theta_3$, $2\theta_4$......... The photographic film records the intercepts of these cones.

From the measured 2θ values, known λ value, and properly sorted n values, (see example 22.11) we deduce the various lattice spacings (*i.e*, d values) of the given crystal. Thus *powder method is used for determining the various d-values for crystal of a given material, and hence in crystal structure analysis*. In contrast the Bragg method is used for determining λ from known d (say d_{100}). Hence it is used in *X*-ray spectroscopy.

The large angle beam (*e.g.*, P_1 in fig. 22·15) is often called "*back reflection*". Its one advantage is in the accurate determinations of small *changes* in d, with (say) temperature.

Example 22·11. X-rays of $\lambda=0{\cdot}741$ A are incident on a powder specimen. A plane film placed beyond the powder at distance 20·0 cm shows the smallest concentric rings of radii 7·22. 10·74, 17·06 and 32·50 cm. Interpret the observations regarding crystal structure.

Solution. Proceeding as in Ex. 22·9 we deduce in succession tan 2θ, θ and sin θ for the four cases

$$\sin\theta = 0.1723,\ 0{\cdot}2439,\ 0.3458,\ 0{\cdot}4879$$

We note: that the last two value of sin θ are very nearly two times the first two values. Hence it is *reasonable* to take the first two rings as the first order ($n=1$) images and the last two as the second-order ($n=2$) images.

From Bragg equation the two d values are related by

$$\frac{d_1}{d_2}=\frac{\sin\theta_2}{\sin\theta_1}=\frac{{\cdot}2439}{{\cdot}1723}=1{\cdot}416$$

This is close to $\sqrt{2}$ and we therefore conclude that $d_1=d_2\sqrt{2}$. In the cubic system the crystal could have sc or bcc arrangement, but not fcc (see § 22·3)

- Largest d corresponds to largest density per unit area of the plane, and hence to largest prominence.

From the given λ, get from eq. (22·7)

$$d_1=\frac{0{\cdot}741}{2\times{\cdot}1723}=2{\cdot}15\ A\ ;\qquad d_2=\frac{{\cdot}741}{2\times{\cdot}2439}=1{\cdot}52\ A$$

The Laue Method. It uses a pencil of '*white*' X-rays, falling on a *single-crystal* (Fig. 22·17) one of whose symmetry axes is kept parallel to the pencil. On a photographic film kept normal to the beam at some distance we get several spots (not rings like Fig. 22·20). These are called Laue spots.

Let rotation of Fig. 22·17 (a) about the incident beam axis be represented by ϕ. Then, in a general ϕ case, none of the lattice planes of the crystal may have its normal in the figure plane, and hence no scattered beam (like P) may occur. Only for some discrete set of ϕ, like $\phi_1, \phi_2, \phi_3 \cdots \phi_i$, the crystal will have some lattice planes whose normals lie in the figure plane and hence allow some Bragg reflection to be observed. The angle where P occurs will be $2\theta_i$, where θ_i, is the glancing angle for the concerned effective lattice plane. With d_i and θ_i both fixed, the system chooses the fitting λ_i from the white X-rays. Fig, 22·17 (b) shows one such P_i. Thus one gets several spots at a discrete set of ϕ_i values and in each case at a discrete set of distances r_i which lead to θ_i since $r_1 = f \tan 2\theta_i$.

Since each spot in Laue pattern belongs to a different λ_i (which we do not know), it will appear that this method would neither measure λ values nor d values. But getting ϕ dependence on record, in addiiion to θ dependence, makes the method more useful in analysis of crystal structure. We cannot give the complex details here.

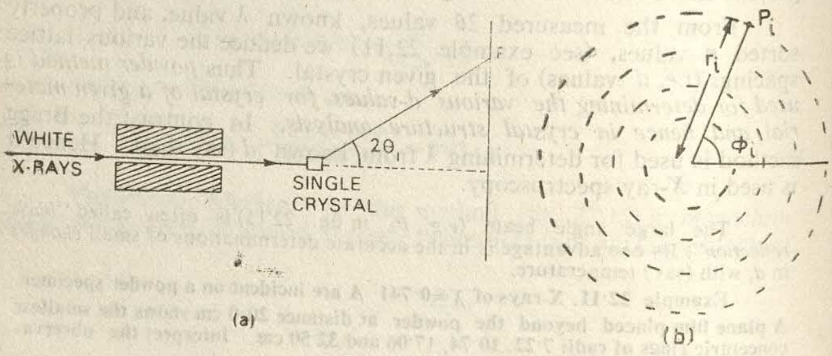


Fig. 22·17.
[*a*] Laue setting [*b*] Laue spots

22·5. Diffraction of Material Particles. In the year 1925 Davisson and Germer reported the results of an experiment on the scattering of a beam of electrons impinging on a single crystal of nickel. (Fig. 22·18). In high vacuum, a beam of electrons is accelerated by a potential of $\sim$50 volts. The beam, narrowed by two diaphragms, falls on the single crystal of nickel normal to the {1 1 1} face. Many of the electrons flow into the crystal, while others bounce back. The intensity of electrons bouncing in different directions is measured by a small collector, whose angular position can be varied. In the collector a retarding potential just short of the accelerating voltage is applied so that only those electrons which bounce with full initial energy are recorded.

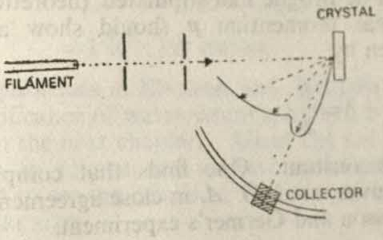


Fig. 22·18.

Davisson-Germer experiment. The electron currents scattered in various directions are shown by broken lines. The curved solid line represents the distribution

The unexpected result of the experiment is that the scattered electrons show a pattern with a well-defined maximum and minimum, resembling a diffraction pattern. When the accelerating potential is varied, the maximum appears most distinctly at 54 volts and is at angle 50° from the incident beam of electrons (see fiig. 22·19).

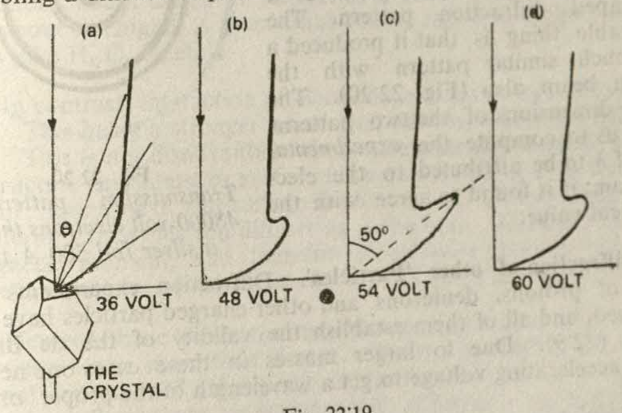


Fig. 22·19.

Electron-diffraction pattern for different accelerating voltages

To explain this, let us tentatively attribute a wavelength λ to the electron beam. (The significance of this will be treated in the next Chapter). As discussed for Bragg reflection, the crystal planes responsible for the maximum must be equally inclined with the directions of incidence and observation. This gives the grazing angle as $\theta=90°-\frac{1}{2}(50°)=65°$. The nickel crystal had a prominent lattice plane actully at this angle from the beam. Further the spacing was $d=0{\cdot}91$ *A*, and using Bragg's equation, the wavelength would be (assuming first order).

$$\lambda=2\times0.91\times{\cdot}9063=1{\cdot}65\ A$$

In 1923 Louis de Broglie had stipulated theoretically that material particles of linear momentum p should show a wave-like behaviour with λ given by

$$\lambda = \frac{h}{p} \qquad \ldots(22{\cdot}9)$$

where h is Planck's constant. One finds that computation for the 54-volt electron gives $\lambda = 1{\cdot}67$ *A*, in close agreement with the value attributed by Davisson and Germer's experiment.

G. P. Thomson in 1927 observed electron diffraction by *transmission* of a fast electron beam through a thin foil of gold. A narrow pencil of *electrons* accelerated through 600,000 volt was sent through a gold film so thin that most of the electrons passed out. Now the 'film' is really composed of numerous very small crystals of gold, so that it acts like a 'powder' and with X-rays it produces a ring-shaped diffraction pattern. The remarkable thing is that it produced a very much similar pattern with the electron beam also (Fig. 22·20). The relative dimensions of the two patterns enable us to compute the *experimental* value of λ to be attributed to the electron beam; it is found to agree with the theoretical value.

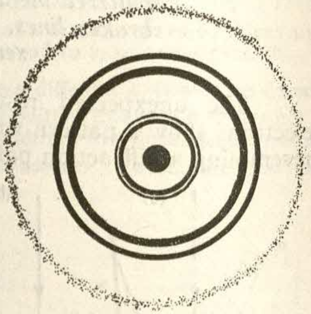


Fig. 22·20.
Transmission pattern of 48000-volt electrons through a silver foil 500 A thick

Diffraction of other 'Particles'. Diffraction experiments with beams of protons, deuterons and other charged particles have been performed, and all of them establish the validity of the de Broglie relation (22·9). Due to larger masses in these cases one needs a smaller accelerating voltage to get a wavelength of the proper order.

Neutral particles like *neutrons* also show diffraction in the same manner. A velocity-selector sorts out neutrons of a narrow range of momentum p, and special detectors have to be used to observe them.

Example 22·12. The smallest angle for strong reflection of a beam of neutrons with a family of crystallographic planes of spacing 3·84 *A* is 30°. Calculate the wavelength of the neutron beam and also the speed of neutrons.

Solution. From the Bragg equation $2d \sin \theta = n\lambda$.

$$2 \times 3{\cdot}84 \times 10^{-8} \times 0{\cdot}500 = 1{\cdot}\lambda$$

We put $n=1$ since the given angle is the smallest,

$$\therefore \quad \lambda = 3{\cdot}84 \times 10^{-8} \text{ em}$$

Now, by de Broglie relation

$$\lambda = \frac{h}{p} = \frac{h}{mv}$$

$$\therefore \quad v=\frac{h}{\lambda m}=\frac{6{\cdot}63\times10^{-27}}{3{\cdot}84\times10^{-8}\times1{\cdot}67\times10^{-24}}$$

$$=\mathbf{1{\cdot}03\times10^{5}\ cm/sec.}$$

⊙

22·6. Applications of Electron and Neutron Diffraction. The theoretical significance of wave-nature exhibited by material particles will be treated in the next chapter. About the use of electron beams and ion beams in microscopy we have spoken in § 15·8. They are now widely used to see the details of the order of a few angstroms—close to molecular sizes.

Here we will discuss the use of electron diffraction and neutron diffraction for studying the crystalline structure of materials. The procedure is the same as with X-rays, but fortunately these methods have complementary character. Let us see this.

X-ray interaction occurs through electromagnetic nature of light acting on the Z electrons of the atom. This interaction is weak ; therefore the X-rays penetrate deep, and X-ray diffraction gives information about the entire interior of a solid. The information is more about the higher Z atoms (like Pb, V, etc) than about the lighter atoms (like H, C, O, etc).

In contrast, interaction of electrons occurs through electrostatic force. This being a stronger interaction the beam does not penetrate deep. This is not disadvantage always, because if you want to study the structure thin films, or surface films on a solid, or even the structure of liquid near the surface (which can be different from that in the interior), then electron diffraction is the best. Electrons also have a magnetic moment, and therefore its behavior also depends on the magnetic character of the surface layers. That is a great advantage.

Neutrons have no charge and their penetration in the solid is very deep—even more than X-rays. But the interaction of neutrons does not depend on Z ; it depends on the fact that it has a magnetic moment μ. So if there are isotopes with different μ, or neighbouring Z atoms with different μ, then neutron scattering experiments will be able to distinguish between them. Another fact is that, for a given interaction strength, the momentum sharing is more in collisions with particles of comparable mass (like H, D, Be, He etc.) than with heavier ones. For this reason neutron diffraction gives far more information about the positions of the lighter atoms in the crystal lattice than what X-rays diffraction does.

Thus for material in bulk X-ray diffraction gives best information for high Z atoms and ignores the effect of μ of the target atoms/ions. Neutron diffraction gives information dominantly depending on μ of the target ions/atoms, and more so for the lightest atoms. Electron diffraction gives no information about the bulk material, but about thin films and surface layers it is the only method. The complimentary features make a combination so useful.

PROBLEMS

22·1. Gold has atomic weight 197 and density 19·3 gm/cc. Deduce an order-of-magnitude value of spacing between the atoms in solid gold. (Given $N=6{\cdot}0\times10^{23}$ per gm mole).

22·2. Distinguish between cubic *I* and fcc lattices. Compute the number of atoms of one kind in a unit cell in each case.

22·3. In a *CsCl* structure the positions of *Cs* atoms are $n_1a\mathbf{i}+n_2a\mathbf{j}+n_3a\mathbf{k}$. Write down the positions of the *Cl* atoms, and the translation vector from any *Cs* atom to any *Cl* atom, or vice versa.

22·4. Show that for the primitive cell of fcc lattice the edges are $a\sqrt{2}$ each with angles 60° between the concurrent edges. Also deduce the volume per primitive cell ?

22·5. Define 'packing fraction' and show that its value is $\pi\sqrt{3}/16$ for the diamond structure and $\pi\sqrt{3}/8$ for an elemental solid in bcc lattice.

22·6. *CsCl* has density 3.98 g/cc and molecular weight 90·5. It has a cubic *P* lattice with Cs^+ at $[\frac{1}{2}\ \frac{1}{2}\ \frac{1}{2}]$ positions relative to Cl^-. Compute (*i*) the lattice constant *a*, (*ii*) the coordination number, (*iii*) the nearest neighbour distance.

22·7. If *a*, *b*, *c* are the primitives along orthogonal axes, show that the spacing between consecutive lattice planes (*hkl*) is given by

$$d_{hkl}=\left[\frac{h^2}{a^2}+\frac{k^2}{b^2}+\frac{l^2}{c^2}\right]^{-1/2}$$

22·8. *NaCl* has fcc lattice with $a=5{\cdot}63$A°. Deduce the spacings of {1 0 0} and {1 1 0} planes. Also compute the dimensions and shape of the primitive cell.

22·9. *KCl* has the same crystal structure as *NaCl* and both are ionic. But for reflections by {1 1 1} planes the odd order spectra are absent in *KCl* while they are only weak in *NaCl*. Discuss this. The atomic numbers of *K*, *Na* and Cl are 19, 11 and 17 respectivety.

22·10. In reflection of ordinary light the only condition required is $\theta'=\theta$, whereas in Bragg reflection an additional condition required is $2d\sin\theta=n\lambda$. Discuss this and also other differences between reflection of light from a *surface* and of X-rays from a *lattice*.

22·11. In a Bragg spectrometer maxima are observed at angles of deviation 11°56′, 24°02′ and 36°24′ from the direct beam. The incident X-rays are monochromatic and of $\lambda=0{\cdot}612$ *A*. Deduce the conclusions about these maxima.

22·12. The powder of a material which has cubic *P* lattice is examined with monochromatic X-rays, using a cylindrical film to record spectra all around the specimen. If λ of X-rays is 0·812 *A* and $a=1{\cdot}920$ *A*, compute the angular positions of the spectra of all the orders due to the lattice planes {1 0 0}, {1 1 1}, and {2 1 1}.

22·13. Define atomic structure factor. Explain why its consideration is significant when λ is of the order of atomic dimensions (or smaller), but insignificant for large λ.

22·14. Define crystal structure factor. Show that for *CsCl* the odd order maxima from (1 0 0) planes will be weak on this account.

22·15. The *X*-ray powder pattern of a simple solid of the cubic system is obtained on a plane film 20·0 cm away from the specimen. The radii of the first few rings are found to be 6·95, 10·28, 16·20, 22·2, and 29·7 cm. Deduce the Bragg angles and find whether the lattice is sc, bcc or fcc.

22·16. Show that the de Broglie wavelength of an electron accelerated to V volt is given by

$$\lambda = \frac{12{\cdot}27}{V^{1/2}}\,\text{A}^{\circ}$$

if *V* is not too large. By what factor would the number 12·27 change if the particle were a proton ?

22·17. Deduce the de Broglie wavelength corresponding to (*i*) 54-volt electron, (*ii*) 5400-volt electron, (*iii*) 0.05 *eV* neutron.

22·18. Show that neutrons of speed $1{\cdot}635 \times 10^{2}$ cm/sec have de Brogile wavelength $0{\cdot}600 \times 10^{-8}$ cm. Deduce the kinetic energy of these neutrons in *eV* unit.

22.19. Which of the three beams—*X*-rays, neutrons, electrons—is best suited to study the (*i*) structure of surface layers, (*ii*) positions of the lighter atoms, (*iii*) positions of heavier atoms, and (*iv*) positions of magnetic ions. Explain the reason in each case.

22·20. A crystal involving equal number of atoms *A* and *B* has *CsCl* structure in the ordered system but in the disordered system *A* and *B* occupy the sites randomly. *A* and *B* have nearly equal atomic numbers, but have different magnetic moments. By what experiment will you decide the extent of ordering ? Expiain.

23

Elements of Quantum Mechanics

Planck *assumed* that a harmonic oscillator could have energy in integral multiples of $h\nu$ only (§ 19·1). Bohr *assumed* that the angular momentum of electrons orbiting in an atom could be in integral multiples of $h/2\pi$ only (§ 20·3). These assumptions violated newtonian mechanics and electrodynamics, but led to satisfactory explanation of many experimental facts. Yet they did not present a proper framework of dynamics for the atomic world.

While the distribution of light past apertures and obstacles could be explained only by attributing wave nature to it, Einstein had to *assume* that in the interaction of light with matter the energy of light showed up in bundles (photons) of energy $h\nu$ only (§ 19·2). This is called the *wave-and-particle duality of light.*

Before this duality could be resolved, Louis de Broglie in 1923 predicted theoretically that material particles should also show interference and diffraction effects as waves do. Thus *matter also showed wave-and-particle duality.*

In this chapter we shall continue to describe a few more developments and experiments on these lines. Towards the end we will briefly introduce the reader to quantum mechanics which now replaces the netwtonian mechanics, though for macroscopic behaviour Newton's laws are very close *approximations* to the results of quantum mechanics.

23·1. De Broglie Waves. In the case of light a bridge between its *wave-character* and *photon-character* can be made thus :

(*i*) Assume that energy and momentum are carried in bundles called photons.

(*ii*) Assume that the travel of the photons is governed by some kind of waves of definite frequency ν.

(*iii*) If the phase velocity of these waves its u, then we have a wavelength defined by $\lambda = u/\nu$.

(*iv*) If the phase velocity depends on λ, then the photon travels with group velocity $c_g = u - \lambda \partial u/\partial\lambda$ (§ 9·7).

Louis de Broglie considered that material particles may also be governed by *phase-waves*, like (iii) above. The frequency to be associated with these waves may be set out through

$$h\nu = \tfrac{1}{2}mv^2 + V = W \qquad \text{...(23·1)}$$

where V is potential energy, $\tfrac{1}{2}mv^2$ the kinetic energy, and W the total energy.

Nothing is said so far about the phase velocity u. But group velocity is related with it through.

$$c_g = u - \lambda\frac{\partial u}{\partial\lambda} = -\lambda^2\frac{\partial}{\partial\lambda}\left(\frac{u}{\lambda}\right) = -\lambda^2\frac{\partial\nu}{\partial\lambda} \qquad \text{...(23·2)}$$

De Broglie now proposed [like (iv) above] that material particles have particle velocity v given by c_g. That gives

$$v = c_g = -\lambda^2\frac{\partial\nu}{\partial\lambda} \qquad \text{...(23·3)}$$

Note that phase velocity u is not specified, but Eq. (23·2) shifts attention to $d\nu/\partial\lambda$, and hence to λ, since ν is defined by Eq. (23·1). That equation, with V held constant, gives

$$\frac{\partial\nu}{\partial\lambda} = \frac{mv}{h}\frac{\partial v}{\partial\lambda} \qquad \text{...(23·4)}$$

Substituting in Eq. (23·3) gives

$$\frac{\partial v}{\partial\lambda} = -\frac{h}{m}\frac{1}{\lambda^2} \quad \Rightarrow \quad v = \frac{h}{m}\frac{1}{\lambda}$$

the integration constant being set as zero. This gives the famous de Broglie relation

$$\lambda = \frac{h}{mv} = \frac{h}{p} \qquad \text{...(23·5)}$$

The waves introduced by de Broglie are variously called pilot waves, 'matter waves', etc. But 'phase waves' describe their function best, since one neither means that matter has a wavy shape. not that matter travels in any direction with a wavy motion ; one only means that when a stream of particles of momentum p passes near an obstacle, or through an aperture, or through a grating, the distribution of a large number of such particles may be explained by considering phase waves associated with them and having a wavelength given by eq. (23·5).

The essence of the aforesaid derivation is that if we have to keep consistent symmetry with the case of optics, then an introduction of $h\nu = W$ (eq. 23·1) *necessarily* leads to $\lambda = h/p$ (eq. 23·5), if the particle is assumed to move with the group velocity c_g.

One may be tempted to consider that eq. (23·5) could directly be deduced from Compton's equation $p=h/\lambda$ for a photon. This is *not* the position. Compton's equation follows from $p=E/c=h\nu/c=h/\lambda$. In the present case the equivalent of $p=E/c$ is absent. All photons, without exception, move with the same speed c (in vacuum), whereas particles of even a given kind (like an electron) may move with different speeds and have correspondingly different momenta. Hence de Broglie's hypothesis is an independent postulate. ∎

We have yet not stated how much u will be. To examine this let us consider the refraction of matter waves. In Fig. 23·1 electrons from an electron-gun G enter a metal box A with speed v_A, and then pass on to another metal box B, the potential of which is different. The electron beam therefore deflects as shown. Treating electrons as particles we note that the lateral velocity does not change. Hence $v_A \sin\theta_A = v_B \sin\theta_B$

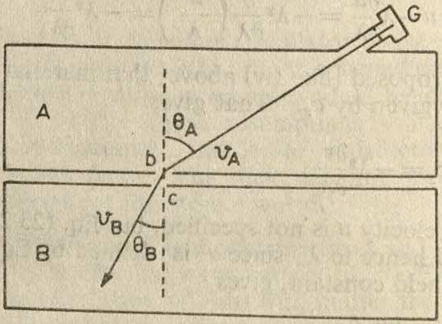


Fig. 23·1
Showing "refraction" of a beam of electrons

$$\therefore \quad \frac{\sin\theta_A}{\sin\theta_B}=\frac{v_B}{v_A} \qquad \text{...(23·6)}$$

On the wave concept, we must have (as in Eq. 10·6)

$$\frac{\sin\theta_A}{\sin\theta_B}=\frac{u_A}{u_B} \qquad \text{...(23.7)}$$

Thus the *phase velocity* u must be inversely proportional to the *particle speed* v, or

$$u=\frac{b}{v} \qquad \text{...(23·8)}$$

where b is a constant during the motion. For light $u=v=c$, and this would set $b=c^2$.

Example 23·1. Make an order of magnitude calculation of the de Broglie wavelength of a 100-volt electron (*i e.*, an electron accelerated through 100 volt potential difference).

Solution. The momentum p is given by the relation $p^2=2mE$,

$$p=(2\times9{\cdot}1\times10^{-31}\times1{\cdot}6\times10^{-19}\times100)^{1/2}$$
$$\approx 5\times10^{-24}\ \text{kg ms}^{-1}$$

Hence, de Broglie wavelength is

$$\lambda=\frac{h}{p}=\frac{6.6\times10^{-34}}{5\times10^{-24}}\approx1.3\times10^{-10}\text{ m.}$$

Example 23.2. Deduce the de Broglie wavelength of (i) neutrons of kinetic energy 1 ev, (ii) thermal neutrons.*

Solution.

(i) Since $p=\sqrt{2mE}$, we get (using 1 ev$=1.6\times10^{-19}$ joule)

$$\lambda=\frac{h}{\sqrt{2mE}}=\frac{6.6\times10^{-34}}{(2\times1.67\times10^{-27}\times1.6\times10^{-19})^{1/2}}$$

$$=2\times10^{-11}\text{ m.}$$

(ii) For *thermal neutrons, energy* $E\simeq kT$. We get

$$E\simeq1.38\times10^{-23}\times300\simeq4\times10^{-21}\text{ joule}$$

$$\lambda=\frac{6.6\times10^{-34}}{(2\times1.67\times10^{-27}\times4\times10^{-21})^{1/2}}=2\times10^{-10}\text{ m.}$$ ⊙

23.2. Direct Evidence for Wave Behaviour of Particles. In § 22.5 we have already cited the evidence in support of de Broglie's hypothesis. We now proceed to describe, two more direct experiments parallel with the biprism interference experiment and the straight edge diffraction experiment of light.

(a) "Biprism" Interference Experiment with an Electron Beam. Fig. 23.2 shows an outline of the arrangement used by H. Ducker in 1955. Here F is a straight hot filament perpendicular to the plane of paper, and A is a cylindrical anode with a wide slit in the front. The heated filament and 20,000 volt accelerating potential between F and A form the *electron-gun*. It gives monoenergetic electrons of de Broglie wavelength$\approx10^{-11}$m (check it).

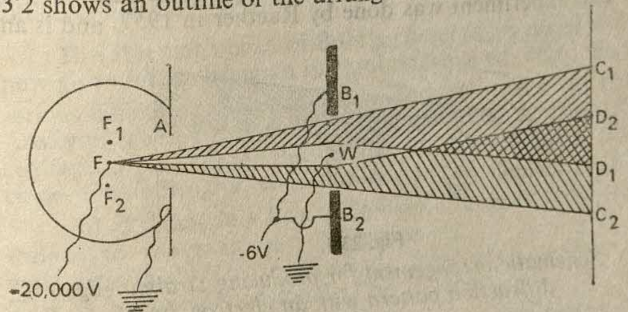


Fig. 23.2.

Schematic arrangement for producing biprism type interference fringes with an electron beam

B_1, B_2 are two metal plates of same thickness and W is a very thin silvered quartz fibre, parallel to the filament F. This wire is given a positive potential +6 volt with respect to the plates B_1, B_2. So there is an electric field between W and B_1, as also between W and B_2. The electric field is largest near W and decreases linearly as we go towards B_1 or B_2. These fields are such that the electron beam passing through the upper half is deflected downwards and meets the screen over area C_1D_1 as if it were coming from F_1 ; the beam passing through the lower half meets the screen over area C_2D_2 as if it were coming from F_2. The area of overlap is D_1D_2. One may see the similarity with a biprism divice giving F_1 and F_2 as images of F by two parts of the biprism.

---

*An assembly of neutrons in equilibrium at temperature T will contain neutrons of means kinetic energy (3/2) kT. In order of magnitude estimates we take $E\approx kT$, and T is taken as $=300K$.

The entire apparatus is enclosed in a high vacuum chamber, so that electrons travel from the filament to the screen without collisions in the path. A photographic record, when seen highly magnified, shows *exactly* the same pattern as with a biprism experiment in light.

One might ask whether slower electrons would not be better. A reduction of accelerating voltage by a factor 100 would increase λ by a factor of 10. As against this advantage of 10-fold larger λ, the electrostatic 'biprism' would need a smaller potential by a factor of 100 (for same separation F_1F_2). It would be extremely hard to ensure stability of such a low voltage and freedom from stray field effects. For these reasons, slower electrons are not preferable for the experiment.

(d) "Straight Edge" Diffraction Experiment with an Electron Beam. The experiment was done by Raether in 1957, and is another

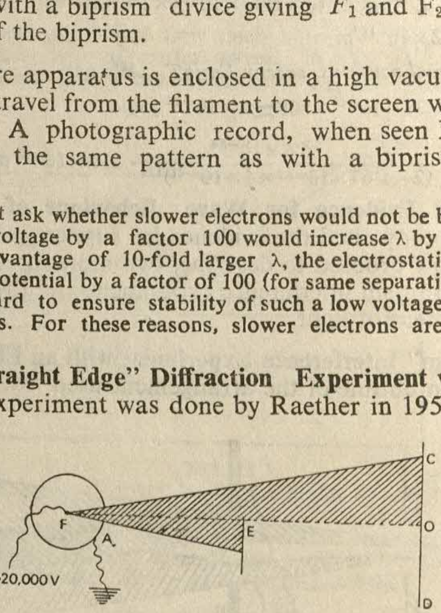


Fig. 23'3.

Schematic arrangement for producing straight edge diffraction pattern with an electron beam

direct parallel with an optical experiment. Fig. 23'3 shows the arrangement. The electron gun is the same as in Ducker's experiments and E is the edge of a single crystal,[*] perpendicular to the plane of paper.

The beam of electrons is received on a photographic film CD, the entire apparatus being kept enclosed in a high vacuum chamber. The pattern abained (when magnified suitably) is found to be indistinguishable from the pattern with light (Fig. 13.11). The wavelength of the electron beam, deduced from the pattern, matches the value calculated from de Broglie's prediction.

[*] The de Broglie wavelength for 20,000 volt electrons is $\sim 10^{-1}$ A, which is $\sim$50,000 times smaller than the wavelength of visible light. Hence the "straightness" of the edge E has to be extremely precise. Crystal edges alone are, therefore, proper. Raether used MgO crystal.

23·3. Some Applications of de Broglie Relation. We shall now apply de Broglie's relation to see how quantized energy levels may arise in a bound system. In a sonometer wire the vibration frequencies are quantized through the standing wave conditions. If we accept that *stationary wave solutions are the only acceptable states in a bound system* then quantization follows naturally from the de Broglie relation. We shall apply it to two cases :

(a) The hydrogen atom. Fig. 23·4 shows schematically the de Broglie standing waves in hydrogen atom. Only those orbits for which the total length of the orbits ($2\pi r$) is an integral multiple of the de Broglie wavelength (λ) are allowed. This gives, for the n^{th} orbit.

$$2\pi r_n = n\lambda_n = n\frac{h}{p_n} \qquad \ldots(23{\cdot}9)$$

where p_n is linear momentum of the electron in the n^{th} orbit. Hence the angular momentum p_ϕ is given by

$$p_\phi = p_n r_n = n\frac{h}{2\pi} \qquad \ldots(23{\cdot}10)$$

which is the same as the Bohr quantization condition.

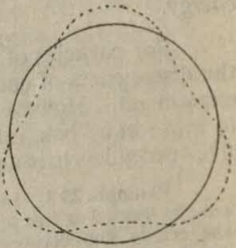


Fig. 23·4.
Schematic of de Broglie standing waves in hydrogen atom [$n=3$]

This is a step ahead of Bohr's postulate, because some '*reason*' is now given for p_ϕ being an integral multiple of $h/2\pi$. But there is one serious difficulty : How do the de Broglie 'waves' go round a circle? A way of looking at it could be that electrostatic force providing the centrepetal acceleration decides the circular motion, and out of all circles the de Broglice condition selects the stationary ones.

(b) Particle in a Box. Consider a particle of mass m constrained to move in one dimension x, with rigid walls* at $x=0$ and $x=a$. This is a one-dimensional "box" of length a.

For standing de Broglie waves in this case, the ends $x=0$ and $x=a$ must be nodes. This condition allows only certain discrete values of the de Broglie wavelength λ *viz*.

$$\lambda_n = \frac{2a}{n}$$

$$\therefore \qquad p_n = \frac{h}{\lambda_n} = \frac{nh}{2a}$$

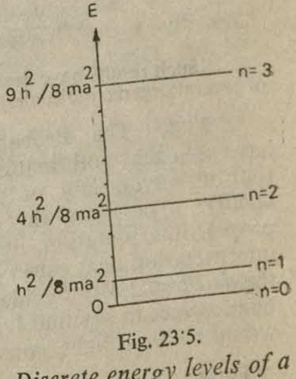


Fig. 23·5.
Discrete energy levels of a "particle in a box"

---

* Rigorously it means that potential energy U is infinite for $x<0$ and $x>a$. In the region $x=0$ to a the potential energy is constant, taken as the zero level.

The energy, which is totally kinetic in this case, is therefore given by

$$E_n=\frac{p_n^2}{2m}=\frac{n^2h^2}{8ma^2} \qquad \text{(with } n=1, 2\ldots\dagger\text{)} \ldots(23{\cdot}11)$$

According to classical mechanics the particle can remain in the "box" with any kinetic energy value, including zero. But according to the present theory the energy states are *discrete*, and the lowest energy is not zero, it is $h^2/(8ma^2)$. The latter is called the *zero-point energy.*

For particles of ordinary mass (m) and ordinary dimensions (a) the discreteness of energy states is too small to be detected in any experiment. However, for particles like an electron constrained to move in a "box" of atomic dimensions the spacing between successive energy levels is appreciable.

Example 23·3. In a long chain organic molecule of length 5A° electrons may be treated as free to move along the length. Deduce the zero point energy, the energy gap between the first two energy states of the electron, and also the wavelength of absorption line arising from this transition.

Solution.

$$E_1=\frac{h^2}{8ma^2}=\frac{(6{\cdot}6\times10^{-34})^2}{8\times9{\cdot}1\times10^{-31}\times(5\times10^{-10})^2}=2{\cdot}4\times10^{-19}\text{J}$$

This is the zero-point energy. Its value *per mole* comes to $NE_1=1.5\times10^5$ J/gm mole.

$$E_2=4.E_1=9{\cdot}6\times10^{-19}\text{ J}$$

$$\therefore\ \text{Energy gap } \triangle E=7{\cdot}2\times10^{-19}\text{ J}=\frac{7{\cdot}2\times10^{-19}}{1{\cdot}6\times10^{-19}}\text{ ev}=4{\cdot}5\text{ ev}.$$

Absorption wavelength λ is given by

$$\triangle E=h\nu=hc/\lambda$$

$$\lambda=\frac{hc}{\triangle E}=\frac{6{\cdot}6\times10^{-34}\times3{\cdot}0\times10^8}{7{\cdot}2\times10^{-19}}\text{ m}=2{\cdot}8\times10^4\ A^\circ$$

[Such results have been checked quantitatively in the absorption spectra of several organic molecules]

23·4. The Probability Interpretation of Phase Waves. We have seen that both matter and radiation show a dual character—that of waves, and of particles. This is called the *wave-particle duality*. The phase waves, introduced by de Broglie, were given a proper interpretation by Max Born in 1926. According to his interpretation light *waves* guide the motion of *quanta*, and similarly the *de Broglie waves* guide the motion of matter *particles*. It would be incorrect to say that light *consists* of waves ; the correct statement would be that light consists of quanta whose arrival anywhere is

---

† For $n=0$, $\lambda_n=\infty$. Such waves cannot have nodes at $x=0$ and $x=a$ unless 'amplitude' is zero. As we shall see later, the square of amplitude of de Broglie waves at any place expresses the 'probability' of the particle being there. So 'zero' amplitude case is not an acceptable solution.

guided by phase waves. Similarly matter consists of particles whose arrival anywhere is guided by phase waves. Thus the concept of waves, on the one hand, and quanta or particles, on the other, are not mutually exclusive but complementary to each other. We may now talk of wave-particle *complementarity* rather than duality.

This should not lead us to the conclusion that light is just a special kind of matter. A striking distinction is given by the fact that light of all kinds travels in vacuum always with the speed 3×10^{10} cm/sec. ; matter never travels that fast*.

The waves predict the distribution of photons or particles under a given setup. Thus when electrons of a certain velocity v are incident on a nickel crystal, we predict their distribution in space by computing the propagation of the waves of wavelength $\lambda=h/mv$. Interpretation of the diffraction pattern now requires the hypothesis that *square of the amplitude of the wave at any point in space gives us the probability of finding electrons travelling through that region of space.* This hypothesis was given by Born.

This probability approach means that if a diffraction pattern is to be observed with a very weak beam sent through an aperture, we may observe individual electrons showing up sometimes here and sometimes there; only by observing the *statistical distribution* of billions of electrons we shall get a check on the prediction given by the wave concept.

Exactly the same discussion would apply for a beam of light passing by an aperture or obstacle. With a very weak beam we should be able to observe individual photons showing up sometimes here and sometimes there.

It will be useful to examine in this context some extreme cases for light waves as well as matter waves :

(*a*) For *light waves of very small wavelength* the theory predicts a diffraction pattern so narrow that the wave character does not readily show up. In contrast, the individual photon carries sufficient energy to readily show up. Hence the particle character can be observed with ease, as in gamma rays.

(*b*) For *light waves of very large wavelength* the individual photons carry so little energy that for any observable effect one would need a flux of a very large number of photons.† Hence the particle character does not show up. In contrast, the wave character predicts pronounced diffraction patterns which are readily observable, as in radio waves.

(*c*) For *very short matter waves* the theory predicts negligible diffraction (rectilinear propagation). This is Newton's first law of motion which holds for all bodies whose mass m is not too small.

---

* For matter this speed is the *upper limit* (see chapter 24).

† The minimum energy flux $\sim10^{-17}$ Js^{-1} needed by a radio receiver amounts to over 10^9 photons/s, assuming $\lambda=30$ *m* (check it).

Wave mechanics would approximate to Newtonian mechanics if $\lambda = h/mv$ is too small compared with the dimensions of obstacles or apertures.

(*d*) For *very long matter waves* one would get diffraction effects predominantly. But such waves do not occur in practice. The lightest particle is the electron and even for that one would need very slow speed to get a de Broglie wavelength comparable with sizes of daily-life objects. Such slow electrons cannot be handled experimentally.

The fact that practically all matter waves are so much shorter in wavelength compared to waves of light explains why diffraction of material particles was observed two and a half centuries later than the diffraction of light.

Born's interpretation of experimental results in terms of *probability* has serious philosophical implications. It would mean the end of newtonian determinism for individual events at the atomic level, and acceptance of a statistical view.

G.I Taylor performed an interesting experiment in this context, which we proceed to describe. Fig. 23·6 shows the standard arrangement for observing diffraction due to a needle : *A* is a source of monochromatic light, *S* is a narrow slit, *N* is a needle placed parallel to the slit, and *P* is a photographic plate, which records the diffraction pattern.

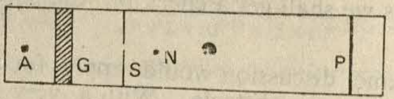


Fig. 23·6.
G.I. Taylor's experiment

G.I. Taylor, while repeating this experiment, added a heavily smoked screen *G* in the path. The parameters were chosen* in such

---

* Let *A* be a 2 watt filament. radiating at efficiency 5% and let the red filter pass 2% light. If slit is at 50 cm distance and has area 10^{-2} cm², the number of photons passed per second in given by

$$N = \frac{2 \times {}^{\cdot}05 \times {}^{\cdot}02 \times 10^{-2}}{4\pi (50)^2 \times 6 \times 10^{-34} \times 3 \times 10^8 / (6 \times 10^{-7})}$$
$$\approx 2 \times 10^9 \text{ photons/sec.}$$

These spread over a length 3×10^8 m/sec, and if the length *S* to *P* is 1 m, the photons in lengh *S* to *P* at a time would be

$$n \approx \frac{2 \times 10^9 \times 1}{3 \times 100^8} \approx 6 \text{ photons}$$

If the smoked screen now has a transmission $\sim 0{\cdot}01$, the photons in the box at any chosen instant would be *zero* most of the time, *one* rarely and *two* almost never.

a way that in the space between S and P there was, on the average, very *much less than* 1 *photon at a time*. This would mean that there would be no multiple photons in space SP to *interfere with each other*. Under these conditions Taylor gave a 3-month long exposure to the film at P, and he found that the pattern recorded was *exactly the same* as that obtained in the absence of the smoked screen G with a much shorter exposure.

This experiment does not permit us to picture the photon as a wavy bundle as we did in Fig. 19·4. If each photon were itself a wave-like bundle, then since no two bundles are present simultaneously in the space SP, no interference should be observed. So we must assume no shape for the photons; they just carry the energy and momentum of the radiation. The phase waves fix the 'probability' of the arrival of a photon at any region of the plate : Wherever the (amplitude)^2 of the wave is high, there is greater probability of arrival of the photons ; and wherever the (amplitude)^2 of the waves is low, there is smaller probability of arrival of the photons.

A corresponding interpretation will apply to the de Broglie waves. The (amplitude)^2 of these waves at any point gives the probability of finding the particle at that point.

Example 23·4 Imagine 5 *Mev* gamma rays passed through a 'slit' of width of the order of atomic spacings in crystals*, say 2*A*. Compute the angular spread. (1 *Mev*=1·6×10^{-13} J).

Solution. For photons,

$$\lambda=\frac{hc}{\text{energy}}=\frac{6{\cdot}6\times10^{-34}\times3\times10^{8}}{5\times1{\cdot}6\times10^{-13}}=2{\cdot}5\times10^{-13}\ \text{m}$$

$$\text{angular spread}\approx\frac{\text{Wavelength}}{\text{aperture}}=\frac{2{\cdot}5\times10^{-13}\ \text{m}}{2\times10^{-10}\ \text{m}}$$

$$=1{\cdot}2\times10^{-3}\ \text{radian.}$$

23·5. Heisenberg's Uncertainty Relations. In § 9·6 we have discussed wave-groups of different lengths (Fig. 9·6) and their Fourier transforms (Fig. 9.5), which together mean that the longer the wave group in x or t the shorter is the range of k (and hence λ) or W (and hence T) values involved in it.

Now, de Broglie waves may also be wave-trains of different lengths. In Fig. 23·7 we represent three different cases. In case (A) the wave-train has infinite length, and hence the wavelength λ is sharply defined ; the uncertainty or spread $\Delta\lambda\rightarrow0$. By de Broglie relation the momentum p of the particle represented by this wave-train is sharply defined. In other words the uncertainty $\Delta p\rightarrow0$. What about the uncertainty Δx in the position of the particle ?

---

* In practice 'slits' of such aperture cannot he had. But the computation gives the order of magnitude of γ-ray diffraction effects in passing through crystals.

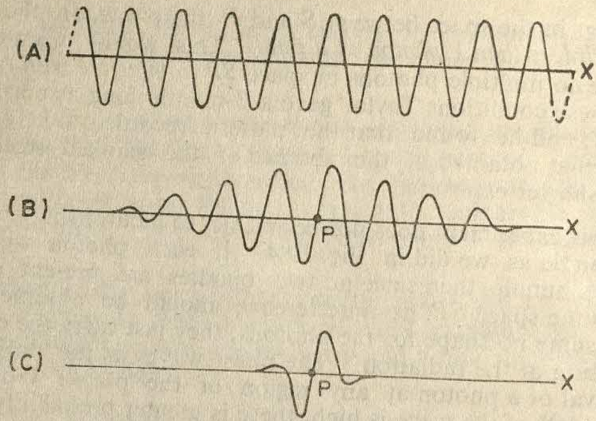


Fig. 23 7.
Wave-groups of different lengths

In case (B) the wave-train is finite, and hence there is some spread $\Delta\lambda$ and consequently some uncertainty Δp in the momentum of the particle represented by this wave. What about Δx for this case ?

In case (C) the wave-train is a pulse, and hence spread $\Delta\lambda$ is very large, *i.e.*, Δp is very large indeed. What about the position uncertainty Δx for the particle ?

Let us now recall Born's interpretation that *square of amplitude of the de Broglie wave at any point gives us the probability of finding the particle at that point*. Then in case (A) the amplitude* is everywhere the same (infinite wave-train) ; so the particle has equal probability of existence anywhere. This makes the uncertainty in position (*i.e.*, Δx) infinite. In case (B), the amplitude and hence the probability is maximum at the centre P of the wave-group and falls off on either side. So the position is now specified within a certain range Δx. Finally, in case (C) the amplitude of the de Broglie wave has a very small spread. Outside a narrow region of x the amplitude is zero everywhere. So the position of the particle is now specified with far less uncertainty Δx.

So, we find that as the momentum uncertainty Δp increases, the position uncertainty Δx decreases, and vice versa. If we make $\Delta p \rightarrow 0$ (truly monochromatic waves), then Δx tends to infinity. If we make $\Delta x \rightarrow 0$ (sharp pulse), Δp tends to infinity. Heisenberg in 1927 expressed this in the form of a specific uncertainty relation : *We cannot specify simultaneously the position and momentum coordinates of a particle with infinite precision.*

---

* The ordinate in Fig 23·7 is the instantaneus ψ value, not the *amplitude* of ψ. Examination over a period T leads to the amplitude (=max ψ).

Heisenberg went a step further. He showed that the product $\triangle p.\triangle x$ had a definite lower limit given by Planck's constant h. We shall not proceed to formally establish this. But an estimate is easy, considering superposition of just two waves of wavelengths λ and $\lambda+\triangle\lambda$. The resultant amplitude becomes zero in a range $\triangle x$ given by

$$\frac{\triangle x}{\lambda}-\frac{\triangle x}{\lambda+\triangle\lambda}\approx 1$$

Using $\lambda=h/p$, this leads to

$$|\triangle p.\triangle x|\approx h$$

More properly Heisenberg puts it as*

$$\triangle p.\triangle x\geqslant h \qquad ...(23{\cdot}12)$$

This is called *Heisenberg inequality or uncertainty relation.*

Acceptance of the "uncertainty" principle is not a defeat of science. A scientist does not *make* a world from his thoughts or axioms. He analyses the laws of nature as they are. "Uncertainty" now appears as one of the laws of nature. The cherished ideal of older science was : "If you give me all the coordinates of all the particles in the universe today, I can predict the entire future precisely". That is called the law of causality. The success of Newtonian mechanices had raised hopes of achieving that. But that is now found to be untenable.

Of course we do measure $<x>$ and $<p>$ of atomic particles and calculate them with definiteness. Here, the symbols $<>$ denote 'expectation value' or 'mean value'. Outside the atomic world there is hardly any distinction between stating x and stating $<x>$ because h is so small and hence $\triangle x$ is too small to matter. But when it comes to atomic particles, all our mechanics has to be so built that the uncertainty principle is satisfied.

There is no meaning in saying that at a given instant the position and momentum of a particle are *exactly x* and *exactly p* respectively. Of course, we cannot prevent any one to *think* of sharp x and sharp p simultaneously, but science goes by measurements and so far as *measurements* are concerned the Heigenberg principle sets the limit for the smallest uncertainties in *simultaneous* measurements of x and p for any particle. It is further accepted that what cannot be measured has no meaning at all in science.

If the horizontal axis in Fig. 23·7 represents time, and $\triangle t$ is length of the wave-train, then the Fourier transform (Fig. 9·5) shows a spread $\triangle\nu$ in frequency, which is in inverse proportion to $\triangle t$. Using $\triangle E=h\triangle\nu$, we get

$$\triangle E.\triangle t\geqslant h \qquad ...(23{\cdot}13)$$

---

* $h/2\pi$ or $h/4\pi$ is sometimes used on the right hand side of eqs. (23.12). This depends on how the range of uncertainty in x and p is defined.

This is another form of *Heisenberg's uncertainty relation.* This means that if a system can stay in a given state for very long, ($\triangle t$ very large), then the energy of the state is very sharply defined (ΔE very small) and vice versa. In fact the principle applies to several other pairs of measurables, whose product has the dimensions of energy × time, *i.e.*, those of h.

23.6. Some Applications of the Uncertainty Relations. Since the uncertainty relations arise from de Broglie wave concept and relations $\lambda = h/p$ and $W = h\nu$ applied to particles as to radiations, one would expect some limiting results to follow directly from the uncertainty relations. We examine some of them here.

(*a*) **Heisenberg's gamma-ray microscope.** Heisenberg gave this thought experiment.* Consider a particle at P in Fig. 23.8, observed through a microscope, the objective of which is shown. The smallest error $\triangle x$ in the position measurement which the microscope would make is given by the diffraction theory (eq. 15.22) as

$$\triangle x \approx \frac{\lambda}{\sin\theta} \qquad \text{...(23.14}a\text{)}$$

where λ is the wavelength of light used to see the particles.

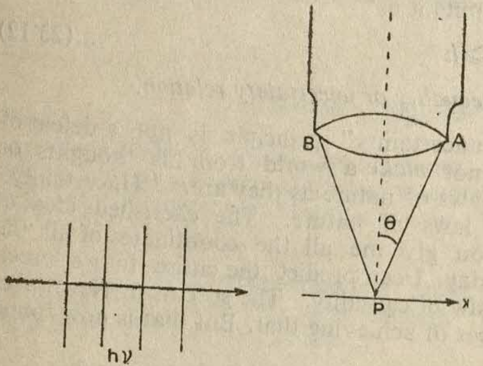


Fig. 23·8.
The gamma-ray microscope

Now to make $\triangle x$ as small as possible, we reduce the wavelength of light used, and go over to γ-rays (in concept at least). But it must be realised that to see the particle *at least one photon* must inter- with it and enter the microscope. This photon of wavelength λ has momentum h/λ. How much of this momentum is shared with the viewed particle is computable *if* the direction of the scattered photon is sharply known. But due to the angular range between PA and PB this computation is uncertain to the extent $(h/\lambda)\sin\theta$, since change in the photon momentum is between $(h/\lambda)(1+\sin\theta)$ and $(h/\lambda)(1-\sin\theta)$, leaving an uncertainty $\pm(h/\lambda)\sin\theta$. Thus the *very process of observation of position* makes the momentum of the electron uncertain by

$$\triangle p_z \sim \frac{h}{\lambda}\sin\theta. \qquad \text{...(23.14 }b\text{)}$$

---

* Though not a real experiment, it is an experiment involving physical measurements. In this sense it is distinct from a philosopher's speculation.

These together lead to (Eq. (23·12). Heisenberg used this method to prove the uncertainty relation, which did not involve Born's interpretation and treatment of a wave-train.

(*b*) **Diffraction at a Slit.** We have seen in diffraction theory that a wave of wavelength λ passing though a slit of width a shows an angular spread of $\theta=\pm\lambda/a$. We shall now make the statement that what is passing through the slit is a *photon* or a *particle.*

Now consider the uncertainty in the *transverse* location of the particle. This is given by $\triangle y=a$, the width of the opening. As regards, the momentum, it is h/λ along the direction of travel (*i.e.*, p_y is zero) at incidence, but because of diffraction there is an angular spread $\theta=\lambda/a$, and hence in the emergent beam p_y has an uncertainty given by

$$\triangle p_y=\frac{h}{\lambda}\theta=\frac{h}{\lambda}\cdot\frac{\lambda}{a}=\frac{h}{a}$$

Thus we get

$$\triangle p_y\triangle y\approx h \tag{23.15}$$

Notice that the deduction applies as much to a photon as to a particle.

(*c*) **Particle in a box.** Consider a one-dimensional box of length a. We may now treat a as the maximum uncertainty in position, $(\triangle x)_{max}$.

Then the uncertainty relation gives the momentum uncertainty that gives the range of lowest p.

$$(\triangle p)_{min}=h/(\triangle x)_{max}=h/a.$$

We may take p ranging *at least* in the range $+h/2a$ to $-h/2a$. The minimum energy is therefore

$$E_{min}=\frac{1}{2m}\left(\frac{h^2}{2a}\right)^2\approx\frac{h^2}{8ma^2} \qquad \text{...(23.16)}$$

Notice that unlike § 23.2 (*b*), this procedure gives only the lowest (*i.e.*, the zero-point) energy, not the complete discrete set of energies.

(*d*) **Electron position in Bohr orbit.** In the case of rotatory motion the equivalent of eq, (23.8) will be

$$\triangle p_\phi.\triangle\phi\geqslant h \tag{23.17}$$

Now, specifying $n=1, 2,$...amounts to stating $p_\phi=h/2\pi,\ 2h/2\pi...$ Thus any specification of n makes the uncertainty $\Delta p_\phi<h/2\pi$. Substitution in Eq. (23.17) leads to

$$\Delta\phi>2\pi$$

This means that in view of the Heisenberg uncertainty relation the angular position of an electron in a Bohr orbit is totally uncertain. If means that the very process of specifying a discrete orbit does *not* permit us to describe the sequence of angular positions ϕ in the path. Bohr's picture thus cannot be taken literally in view of quantum mechanics.

23.7 The Schrodinger Wave Equation. We have described in § 23.1 the essential equations for 'waves' associated with particles, as introduced by Louis de Broglie in 1923. These define ν and λ for the wave through the equations

$$\nu=\frac{W}{h}=\frac{1}{h}\left(\frac{p^2}{2m}+V\right) \qquad \text{...(23.18)}$$

$$\lambda=\frac{h}{p} \qquad \text{...(23.19)}$$

where W is the total energy, V is the potential energy and p is linear momentum.

It had become obvious around this time of history that newtonian laws of motion needed a drastic revision of total replacement by a suitable law which would explain the phenomena in the ultra-small world of atoms and molecules, and would approximate to newtonian laws for events on the macro-sacle.

Schrodinger and Heisenberg independently achieved success in this respect in 1925. Later on it was proved that the two solutions were equivalent, and we shall here take up Schrodinger's approach only. It will be seen that the solution lay not in modifying the newtonian equation to take account of wave-like properties, but in modifying the wave-equation so that it would reflect the corpuscular propertises too.

For a particle moving in the x-direction we can write the most general equation of the de Broglie plane wave in the experimental form, as in Eq. (9.40a) :

$$\psi=A\exp\left[-2\pi i\left(ix-\frac{x}{\lambda}\right)\right]$$

The meaning of ψ we shall presently see. Substitution for ν and λ gives

$$\psi=A\exp\left[-\frac{2\pi i}{h}(Wt-px)\right] \qquad \text{...(23.20)}$$

Schrodinger searched for that basic equation for which eq. (23.20) is a solution. This equation was found to be*

---

* With W and p treated as constants, substitution from Eq. (23·20) into Eq. (23·21) gives

$$L.H.S.=\left[-\frac{h^2}{8\pi^2 m}\left(-\frac{4\pi^2 p^2}{h^2}\right)+V\right]\psi=\left(\frac{p^2}{2m}+V\right)\psi=W\psi$$

$$R.H.S.=-\frac{h}{2\pi i}\left(-\frac{2\pi i W}{h}\right)\psi=W\psi=L.H.S.$$

$$-\frac{h^2}{8\pi^2 m}\frac{\partial^2\psi}{\partial x^2}+V\psi=-\frac{h}{2\pi i}\frac{\partial\psi}{\partial t} \qquad \text{...(23.21)}$$

The next step was to state that this equation (23.21) is applicable to all cases of one-dimensional motion, the nature of V deciding what solutions come out. (Eq. 23.21) is called the Schrodinger Wave Equation, containing the time, for one dimensional motion of a particle. For three dimensional general motion the equation would be

$$-\frac{h^2}{8\pi^2 m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+V\psi=-\frac{h}{2\pi i}\frac{\partial\psi}{\partial t} \qquad (23.22)$$

Notice that the Schrodinger equation has not been proved or deduced from the fundamental properties of the medium, as we do, for example, in the case of elastic waves on strings or in an elastic medium or for electromagnetic waves (Chapter 8). In the case of 'matter waves'—that is, the phase waves which govern the probability of arrival of the particle at any region of space—we do *not* know anything about what these waves are. All we know is that with ν and λ chosen according to eq. (23.18) and (23·19) the experimental observations can be satisfactorily explained. Hence we wrote Eq. (23·20) and then searched for a differential equation for which this is a solution. The Schrodinger equation is one such differential equation.

A remarkable thing about this equation (23.21) is the occurrence of $i(=\sqrt{-1})$ in it.* The solution (23.20) also contains i. Hence ψ is a complex quantity and can have no physical significance by itself. But if we take $\psi^*\psi$, where ψ^* is the complex conjugate of ψ, we do get a real quantity. From eq. (23.20),

$$\psi^*\psi=A\exp\left[-\frac{2\pi i}{h}(Wt-px)\right]A\exp\left[+\frac{2\pi i}{h}(Wt-px)\right]$$

$$=A^2\text{, a real quantity.}$$

Max Born's postulate (§23.4) about "square of the amplitude", thus refers to A^2, which is equal to $\psi^*\psi$, often written as $|\psi|^2$. We can now state that postulate as follows :

If dv is any volume element of space and ψ is the wave-function in that part of space then the probability of the particle appearing in that part of space is proportional is $\psi^*\psi\ dv$.

The Schrodinger wave function ψ has some other interesting properties. From eq. (23·20) we deduce

$$\left[-\frac{h}{2\pi i}\frac{\partial}{\partial t}\right]\psi=W\psi \qquad \text{...(23.23)}$$

and

$$\left[\frac{h}{2\pi i}\frac{\partial}{\partial x}\right]\psi=p\psi \qquad \text{.. (23.24)}$$

---

* See also Chapter 9, Eqs. (9·11), (9·29), (9·40*a*) etc.

These mean that the mathematical *operators* (given in square brackets) applied on ψ give us the energy W and the linear momentum p of the particle for which the wave equation (23·20) was written out.

The important step now was that Schrodinger postulated eq. (23·22) as the basic law of mechanics governing the motion of *all bodies.* This law replaced Newton's second law of motion. This generalisation by Schrodinger revolutionised scientific thought. We passed on from *causality* (or determinism) to *probability* (or indeterminism) in the deduction of results of an analysis. The experiments in the realm of atoms and photons pointed towards the inevitability for such a change, and Schrodinger found the law which would give results in line with experimental evidence. Newtonian mechanics was thus replaced by *wave mechanics* in the year 1925.

As stated earlier, Heisenberg in the same year had developed a matrix method of obtaining the same results. Since his *matrix mechanics* was proved equivalent to Schrodinger's wave machanics, the name more suitable for the new mechanics is *quantum mechanics.*

The starting point in any problem is now a knowledge of potential energy V as a function of space coordinates.* We substitute this V in Schrodinger's equation (23·22) and search for a solution for ψ. From that the desired information is then extracted through the operators.

The Stationary State Schrodinger Equation. For a system in which the total energy W does not change with time, the right hand side of eq. (23·22) may be replaced by $W\psi$ in view of eq. (23·23). That gives us

$$\frac{h^2}{8\pi^2 m}\left(\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}\right)+(W-V)\psi=0. \qquad ...(23.25)$$

This is called the *stationary state Schrodinger equation* or the *time independent Schrodinger equation.*

An important feature that comes out in solving this equation is that for negative values of W (*i.e.* for bound systems) the equation does not lead to *acceptable solutions* except for some particular values $W_1, W_2 \ldots, W_n$. The condition for acceptability is that ψ should be *single-valued, finite and continuous over all parts of space.* The reason is that ψ leads to specific measurable properties (through suitable operations), and these properties have to be single-valued and finite everywhere.

The particular W values for which acceptable solutions exist are called the *eigenvalues of energy*, and the corresponding ψ functions are called *eigenfunctions* belonging to the corresponding energies. Note that once we accept the Schrodinger equation as the basic law of mechanics, we do not have to apply arbitrary quantum conditions, as had to be done by Planck and by Bohr. The quantum conditions come out naturally, as part of the new mechanics.

---

* In Newtonian mechanics a knowledge of *force* is the starting point.

23·8. Postulatory Basis of Quantum Mechanics. Instead of the historical growth through Planck's $E=nh\nu$, Einstein's $W=h\nu$, Bohr's $p_{\phi}=nh/2\pi$, and de Broglie's $\lambda=h/p$ and then Schrodinger's use of equation (23·22), we may just introduce quantum mechanics through a set of postulates. Note that Newton's laws are regularly taught as postulates, omitting whatever history there was in their growth. In that line it is now becoming customary to initiate quantum mechanics in terms of *postulates*, for which no proofs are needed, except that they lead to a satisfactory and quantitative explanation of experimental observations.

(a) The First Postulate. *Information about the properties of a system of particles is contained in a function* ψ *of all the space coordinates and time.*

This function is called the wave function of the system. In order to be acceptable, the function (and its derivatives) should be single-valued, continuous and finite everywhere, going over to zero at infinity.

(b) The Second Postulate. *The information regarding any physical variable of a system can be extracted from the wave function by the use of the corresponding operator for the variable.*

If f is a physical variable and F its operator the rule is

$$F\psi=f\psi \qquad \text{...(23·26)}$$

The left side means the operator F acting on the function ψ, while the right side means f multiplied by ψ.

The list of operators is given in Table 23·1. For any function of q, p and t the operator will be the corresponding function of the corresponding operators. As three examples :

(*i*) Operator for p_x^2 is given by

$$P_x^2=-\frac{ih}{2\pi}\frac{\partial}{\partial x}\left(-\frac{ih}{2\pi}\frac{\partial}{\partial x}\right)=-\frac{h^2}{4\pi^2}\frac{\partial^2}{\partial x^2} \qquad \text{...(23·27)}$$

TABLE 23·1

The Wave Mechanical Operators

Variable	*Symbol*	*Operator Symbol*	*Operation*
Coordinate	q_k	Q_k	multiply by q_k
Linear momentum	p_k	P_k	$-\frac{ih}{2\pi}\frac{\partial}{\partial q_r}$
Time	t	τ	multiply by t
Total energy	W	E	$+\frac{ih}{2\pi}\frac{\partial}{\partial t}$
Potential energy	V	V	multiply by V

(*ii*) Since x component of angular momentum is given by $(yp_z - zp_y)$, the operator for angular momentum component L_x is

$$L_x = -i\frac{h}{2\pi}\left(y\,\frac{\partial}{\partial z} - z\,\frac{\partial^2}{\partial y}\right) \qquad \ldots(23{\cdot}28)$$

(*iii*) Since kinetic energy is given by $(p_x^2 + p_y^2 + p_z^2)/2m$, the operator for *total* energy is given by

$$H = -\frac{h^2}{8\pi^2 m}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + V \qquad \ldots(23{\cdot}29)$$

This operator H is called the *Hamiltonian operator.*

The Schrodinger wave equation (23·22), as may be seen, comes as a result of identifying this operator H with E of Table 23·1.

(c) The Third Postulate. *The expectation value f or* $<f>$ *of a variable f for a system whose wave function is* ψ *is given by*

$$<f> = \frac{\oint \psi^* F \psi\, dq}{\oint \psi^* \psi\, dq} \qquad \ldots(23{\cdot}30)$$

where ψ^* is the complex conjugate of ψ and integration over dq means integration over all values of all the space coordinates.

If ψ is a solution for Schrodinger equation, $a\psi$ will also be a solution, where a is a simple number. For this reason the amplitude in ψ is often so chosen that the denominator of eq. (23·30) is unity. The wave function is then said to be *normalised.* With a normalised wave function

$$<f> = \oint \psi^* F \psi\, dq \qquad \ldots(23{\cdot}31)$$

(d) The Fourth Postulate. *The probability of a spectral line being observed for transition between two states 1 and 2 defined by wave functions* ψ_1 *and* ψ_2, *is proportional to**

$$R_{12}^2 = [\oint \psi_1^* q \psi_2\, dq]^2 \qquad \ldots(23{\cdot}32)$$

This is just one part of a vast range of applications dealing with the time development of states. Electric dipole moment has value qe and the integral above is an evaluation of the oscillating dipole strength. If R_{12} comes to zero, no dipole radiation corresponding to transition $1 \rightleftharpoons 2$ would exist.

Unlike Bohr's theory, we now have a quantitative way of evaluating the relative probabilities of the different conceivable transitions.

---

* This refers to electric dipole transition. There are also other (much weaker) transitions, which we ignore at present.

In Heisenberg's method the attention is centred on the *observables*. As one example R^2_{12} relates to an observable spectral line, whereas neither ψ_1 nor ψ_2 is an observable. If a direct procedure leads to R^2_{12} one saves not only the *trouble* of going through ψ's, but one also avoids the *confusion* that may arise from trying to draw mental pictures of ψ or $|\psi|^2$ distributions. But since we are trained in the process of drawing pictures (of even non-observable features) the Schrodinger method is taught at earlier levels. At higher levels we find it impossible (and quite unnecessary) to go through this labrynth and directly use the Heisenberg matrix method, where matrices lead to $|R_{ij}|^2$ values and other observable features. Symmetry considerations are of special significance in telling which R_{ij} values must be zero, and which ones may be finite (non-zero). That needs learning of what is called *group theory*, besides *matrix algebra*. We will, however, resist the temptation of talking more about this fascinating field, and restrict ourselves to the Schordinger method.

23·9. Some Applications of Schrodinger's Equation. We will give here some examples for one-dimensional systems involving constant total energy W. The equation for use is [from Eq. (23·25)],

$$\frac{h^2}{8\pi^2 m}\frac{\partial^2\psi}{\partial x^2}+(W-V)\psi=0 \qquad \text{...(23·33)}$$

But the nature of potential function V varies from case to case.

(a) Particle in a One-dimensional Box. We have discussed this case twice earlier. But formal wave mechanics brings in a definite advancement in the information. Let us see.

In this case, at the box boundaries (at $x=0$ and $x=a$) and *outside the box*, $V=\infty$, which leads to $\psi=0$.

For *inside the box* $V=0$ reduces eq. (23·33) to

$$\frac{h^2}{8\pi^2 m}\frac{\partial^2\psi}{\partial x^2}+W\psi=0 \qquad \text{...(23·34)}$$

for which the most general solution would be

$$\psi=A_1\exp(ibx)+A_2\exp(-ibx)\ ; \qquad b^2=\frac{8\pi^2 mW}{h^2}.$$

But the requirement of ψ being single valued leads to $\psi=0$ at $x=0$ and at $x=a$. This is satisfied only if

$$b=\frac{n\pi}{a} \text{ and } A_1=-A_2=\tfrac{1}{2}A_n \text{ (say)}$$

These lead to a discrete set of values for W given by

$$W_n=\left(\frac{h^2}{8ma^2}\right)n^2, \qquad (n=1, 2,...)$$

and corresponding eigenfunctions

$$\psi_n=\tfrac{1}{2}A_n\,[\exp(ibx)-\exp(-ibx)]$$

$$=iA_n\sin\frac{\pi nx}{a}. \qquad \text{...(23·35)}$$

The normalisation condition gives

$$\int_0^a A_n^2 \sin^2\left(\frac{\pi n x}{a}\right) dx = 1 \Rightarrow A_n = \sqrt{\frac{2}{a}}$$

The normalised eigenfunction belonging to energy W_n is thus

$$\psi_n = i\sqrt{\frac{2}{a}} \sin \frac{\pi n x}{a} \qquad \ldots(23{\cdot}36)$$

The probability distribution of finding the particle per unit length at x inside the box is given by

$$\psi_n^* \psi_n = \frac{2}{a} \sin^2 \left(\frac{\pi n x}{a}\right) \qquad \ldots(23{\cdot}37)$$

It will be interesting to plot $\psi_n^*\ \psi_n$ against x for the states. $n=1, 2, 3, 4$ and see the differences. Notice that the amplitude (2/a) is independent of n.

The intensity of radiation for a transition $n \rightarrow m$ is proportional to R_{nm}^2, where

$$R_{nm} = \int_0^a \psi_n^* x\ \psi_m\ dx.$$

This comes out *zero* if both n and m have the same *parity* (both odd or both even), and *non-zero* if out of n and m one is odd and the other even. The value then comes to

$$R_{nm} = \frac{8a}{\pi^2} \frac{nm}{n^2 - m^2} \qquad \ldots(23{\cdot}38)$$

(b) Harmonic Oscillator. For one dimensional oscillator the potential energy is $V = \frac{1}{2}Kx^2$, to be put in Eq. (23·33). The solutions for the case require somewhat complex mathematics and we shall therefore give only the results.

The eigenvalues of energy come out to be

$$W_n = (n + \tfrac{1}{2})h\nu \qquad (n = 0, 1, 2\ldots) \qquad \ldots(23{\cdot}39)$$

where

$$\nu = \frac{1}{2\pi}\sqrt{\frac{K}{m}} \qquad \ldots(23{\cdot}40)$$

Thus ν in Eq. (23·39) equals the classical frequency, but the energy states are quantized. The energy status are not $nh\nu$ but $(n+\frac{1}{2})h\nu$. This is a departure from Planck's *assumption*. But the energy *steps* are of equal size $h\nu$ and hence Planck's results remain unaffected, except for the addition of *zero point energy* $\frac{1}{2}h\nu$ per oscillator. There are many consequences of this, particularly in thermodynamics, which have been duly found correct.

The $\psi_n^*\ \psi_n$ function for the case is best shown graphically. Fig. 23·9 shows this for $n=0$, $n=1$ and $n=8$, where the classical limits of

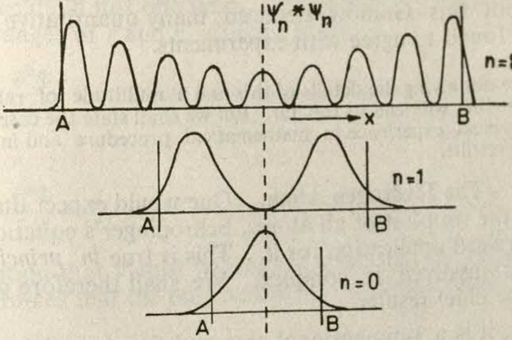


Fig. 23·9.
Probability distribution in three states of the harmonic oscillator (schematic)

x for corresponding energies are shown by *A* and *B* in each case. We note serious departures from classical mechanics in three respects :

(*i*) For each state, the probability function $\psi_n{}^*\psi_n$ has a finite value beyond the classical limits *A* and *B*.

(*ii*) For the lowest state the probability is maximum at the centre, whereas classical theory of oscillations gives that maximum time is spent near the ends.

(*iii*) The probability function does not vary smoothly over x but shows $n+1$ maxima for the n^{th} state.

Each of these predictions has been verified by quantitative experiments, establishing the credentials of wave mechanics.

(c) Tunneling or Barrier Penetration. Let a particle of kinetic energy *E* reach a rectangular potential barrier of finite height V_0 and width *a*. Then classical mechanics would tell that for $E<V_0$ the particle would never cross the barrier. But in wave mechanics when we solve the Schrodinger equation for the three regions (before, inside, and beyond the barrier) we find that $|\psi|^2$ comes out non-zero both inside and beyond the barrier even for $E<V_0$. We can quantitatively deduce the ratio of $|\psi|^2_{beyond}$ to $|\psi^2|_{incident}$ and call this ratio the transmission factor.

That is a rather strange prediction indeed from quantum mechanics. One would ask : how can kinetic energy be negative, as it would be inside the barrier for $E<V_0$? But the fact is that the *observations* can be carried only outside the barrier. So in wave mechanics the above question is not valid—we only talk about the *observables*. Transmission is an observable. Gamow applied the idea to the emission of α particles by heavy nuclei. The α particles *inside* the nucleus were assumed to have energy E above the state when α would be outside the nucleus, but a potential barrier of height

$V_0 > E$ and a certain width existed between the inside and outside states. From this Gamow deduced many quantitative predictions, which were found to agree with experiments.

We are not giving the details of this and a multitude of related applications, because that will lead us too far. But we shall state the case of hydrogen atom as a typical experience in mathematical procedure and in the interpretation of the results.

23·10. The Hydrogen Atom. One would expect that hodrogen atom being the simplest of all atoms, Schrodinger's equation will have a straightforward application for it. This is true *in principle*, but the mathematics involved is complex. We shall therefore give only an outline of the chief results.

Firstly, it is a 3-dimensional case and the potential energy of an electron in the field of a nucleus of charge Ze ($Z=1$ for hydrogen) is given by

$$V = -\frac{Ze^2}{r} \qquad \ldots(23{\cdot}41)$$

It is therefore necessary to put Schrodinger's equation (23·25) in spherical polar coordinates*[*] :

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2\sin^2\theta}\frac{\partial^2\psi}{\partial\phi^2}$$

$$+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right)+\frac{8\pi^2\mu}{h^2}(W-V)\,\psi=0 \qquad \ldots(23{\cdot}42)$$

We try a solution of the form

$$\psi = \Phi.\Theta.R \qquad \ldots(23{\cdot}43)$$

where Φ is a function of ϕ alone, Θ is a function of θ alone, and R is a function of r alone. The process of solving this way is called the *separation of variables.*[†]

Eq. (23·43) would give

$$\frac{\partial\psi}{\partial r}=\frac{\psi}{R}\frac{\partial R}{\partial r}\,; \qquad \frac{\partial^2\psi}{\psi r^2}=\frac{\psi}{R}\frac{\partial^2 R}{\partial r^2} \quad \text{etc.}$$

Substituting in (23·42) leads to

$$\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)+\frac{1}{\Phi\sin^2\theta}\frac{\partial^2\Phi}{\partial\phi^2}$$

$$+\frac{1}{\Theta\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial\Theta}{\partial\theta}\right)+\frac{8\pi^2\mu r^2}{h^2}(W-V)=0 \qquad \ldots(23{\cdot}44)$$

---

[*] μ is used for electron mass here to avoid confusion with quantum number m.

[†] We will see that in the present case it works.

Now, ϕ occurs in just one term. Therefore, if the equation has to hold for all ranges of r and θ, we must have

$$\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\phi^2} \text{ a constant independent of } \phi.$$

Hence we write

$$\frac{1}{\Phi}\frac{\partial^2\Phi}{\partial\phi^2}=f_1(r, \theta, W)=\text{const}=-m^2 \text{ (say)} \qquad \text{...(23·45)}$$

The condition that ψ must be single valued, *i.e.*, repeating with a period 2π, enforces that the only acceptable solutions of Eq. (23·45) are

$$\Phi=Ae^{im\phi}, \text{ (} m \text{ an integer)} \qquad \text{...(23·46)}$$

This lays down a quantum condition related with the ϕ coordinate.

Next, substitution from Eq. (23·45) in (23·44) makes θ appear only in the second and third terms of that equation. If Eq. (23·44) is to hold for all r values, the sum of these two terms must be a constant independent of θ, So we write

$$\frac{1}{\Theta}\frac{1}{\sin\theta}\frac{\theta}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right)-\frac{m^2}{\sin^2\theta}=f_2(r, W)=C_2 \text{ (say)} \quad \text{...(23·47)}$$

Mathematics tells us that this equation has no acceptable solutions except when

$$C_2=--l(l+1), \qquad (l \text{ an integer} > |\, m \,|) \qquad \text{...(23·48)}$$

This lays down another quantum condition related with the θ and ϕ coordinates.

Finally, substitution from (23.48) in Eq. (23·44) leads to

$$\frac{1}{Rr^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)-\frac{l\,(l+1)}{r^2}+\frac{8\pi^2\mu}{h^2}(W-V)=0 \qquad \text{...(23·49)}$$

Thus far we have assumed V to be independent of ϕ and θ, but have not used how it exactly depends on r. Hence the quantum conditions regarding m and l are applicable to all central force cases.

We now proceed to specify the dependence of V on r. For hydrogen and hydrogen-like cases, the coulombian potential is $V=-Ze^2/r$. Substituting in Eq. (23·49) and trying to solve it tells us that acceptable solutions are available in only two situations :

(*t*) for positive value of W, all values lead to an acceptable solution.

(*ii*) for negative value of W (*i.e.*, bound states), acceptable solutions of Eq. (23·49) are available if, and only if,

$$W_n = -\frac{2\pi^2 \mu e^4 Z^2}{n^2 h^2}, \quad (n = \text{integer} > l+1) \qquad \ldots(23{\cdot}50)$$

This lays down the third quantum condition related with the r, θ and ϕ coordinates.

Result (23·50) is the same as eq. (20·11) obtained from Bohr's theory. If we pay attention to the *eigenvalues of energy* alone, the Schrodinger equation tells nothing more than Bohr's theory. But notice that two other quantum numbers l and m appear naturally in Schrodinger's theory. (We referred to these in § 20·6). Further working tells us that l gives us the *angular momentum* L through the relation*

$$L^2 = l(l+1)\frac{h^2}{4\pi^2} \qquad \ldots(23{\cdot}51)$$

and m gives us the component of angular momentum along a specified direction in space :

$$L_z = m.\frac{h}{2\pi} \qquad \ldots(23{\cdot}52)$$

Eq. (23·51) really means that Bohr's value $L = nh/2\pi$ is not valid ; now for $n=3$ (for example), l can have values 0, 1, 2 and accordingly L^2 can be 0, 2 or 6 times $(h/2\pi)^2$ for same $n=3$. It is found that in explaining the spectra, the properties of atoms, the periodic table, etc. this extension is of great value.

In § 20·6 we had mentioned quantum numbers l, m_l, s, and m_s. The first two are the same as designated by l and m here. Electron-spin does not come out as a solution of eq. (23·42) and one has to introduce relativistic considerations to take account of it.

The *eigenfunctions of* eq. (23·49) for different n, l, m combinations give us some more information. Just as one example we show the radial function R for the 2s state ($n=2$, $l=0$) in Fig. 23·10. Now, a volume element between radii r and $r+dr$ is $4\pi r^2 dr$. Therefore, $4\pi^2 (RR^*)\,dr = D dr$ gives the *probability* for finding the electron between distances r and $r+dr$ from the nucleus. D is called the *radial density distribution function*. In Fig. (23·10) this is also given for the 2s state. The curve shows that in this state D is maximum at $r \sim 5{\cdot}5$ Bohr radii, but there is a hump at $r \sim 1$ Bohr radius too. The radial distribution curve for the 2p state ($n=2$, $l=1$) does not show this extra hump. These differences are of great help in understanding the atomic structure and interactions, but we shall not elaborate further.

---

* In § 20·6 (a) we had written $L = lh/2\pi$. Wave machanics shows two differences. We get the eigenvalues of L^2 not of L. (Which means no experiment would measure L). Also $\sqrt{L^2}$ comes to $\sqrt{l(l+1)}\,h/2\pi$, not $lh/2\pi$. Experiment agrees with the wave mechanical value.

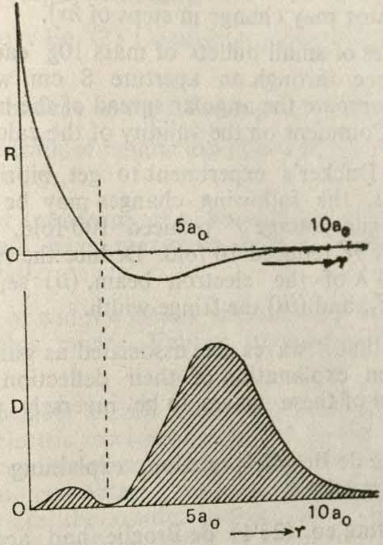


Fig. 23.10.
Plot of radial function R and density distribution function $D=4\pi r^2 RR^*$ for the 2s state

A significant point is that unlike Bohr's *definite* orbits now we talk of only probability distributions. Just as Fig. 23·10 refers to distribution relative to r, one could also show distribution relative to θ or ϕ. This is not the distribution of the electron, but the distribution of the *probability* of existence of the electron in space defined by r, θ and ϕ.

One may get a feeling that we gave up classical laws of physics in favour of quantum laws of Planck, Einstein and Bohr, and now probably these latter laws are also being questioned. That is only partly true. What Planck, Einstein and Bohr introduced were empirical postulates which explained limited experimental facts to a limited extent. A proper theory, applicable in all situations (including the macroscopic ones), was required and quantum mechanics, as developed by Schrodinger, Heisenberg, Born and others fulfils that requirement. In the range of large dimensions and large masses it gives the Newtonian mechanics as the valid approximation, while in the range of atomic dimensions and masses this theory with all its details (and implications) has been found highly successful in explaining the experimental facts and in predicting new things.

PROBLEMS

23·1. For any harmonic oscillator the graph between displacement x and momentum p (or between angular displacement ϕ and angular momentum p_ϕ) will be an ellipse. Prove this. Show that in passing from one ellipse to another, for which the area of p vs. x

ellipse changes by h, the energy change would be $h\nu$, where ν is frequency of the oscillator. [Planck used this logic in establishing that energy of an oscillator may change in steps of $h\nu$].

23.2. A series of small bullets of mass 10g each are fired with speed 2×10^4 cm/sec through an aperture 8 cm wide. Using de Broglie's concept compute the angular spread of the bullets in passing through the slit. Comment on the validity of the calculation.

23·3. In *H.* Ducker's experiment to get biprism type fringes with electron-beams, the following changes may be made, one at a time : (*a*) electron-gun vottage V reduced 100-fold, (*b*) potential of 'biprism' bridge wire W reduced 10 fold. Deduce the resultant changes in each case in (*i*) λ of the electron beam, (*ii*) separation of the virtual sources F_1F_2, and (*iii*) the fringe-width.

23·4. Prove that if waves are associated as guiding the motion of the particles, then explanation of their deflection would require the phase velocity u of these waves to be inversely proportional to the particle speed.

23·5. Deduce de Broglie's relation, explaining the significance of the assumptions involved.

23·6. In writing eq. (23·1) de Broglie had actually used the relativistic energy expression (see chapter 24), which for $v \ll c$ reduces to $m_0c^2+\frac{1}{2}mv^2$. Would this change disturb the logic leading to eq. (23·8) ?

23·7. Use the Bohr condition $mv_nr_n=nh/2\pi$ and stationary wave condition $2\pi r_n=nh_n$ to deduce the expression for de Broglie wavelength.

23·8. An organic long chain molecule shows absorption lines at $\lambda=24{\cdot}0\mu$, $9{\cdot}0\mu$, $4{\cdot}9\mu$, $3{\cdot}0\mu$ $(\mu=10^{-6}m)$. Show that these can be explained by an energy level diagram with $E=n^2\times$a constant, and $n=1, 2, 3, \ldots\ldots$If this is attributed to electrons free to move along the length of the molecule, deduce the length of the molecule.

23·9. Discuss the distinction between these two statements : (1) light consists of waves, (2) light consists of quanta whose arrival at any part of space is guided by phase-waves, with the postulate that probability is proportional to square of the wave amplitude. Take any one experimental situation for the discussion.

23·10. What are the distinguishing features between phase-waves associated with quanta of electromagnetic radiation and phase-waves associated with material particles ?

23·11. Newtonian mechanics breaks down in the realm of atomic phenomena. Comment on this statement. If a new mechanics is developed for atomic phenomena, it should reduce to Newtonian mechanics for bodies of practical sizes. Examine how wave mechanics fits in with this requirement.

23·12. State Heisenberg's uncertainty principle. Discuss its impact on the concept of causality.

23·13. Prove Eq. (23·13) using a procedure similar to that for Eq. (23·12).

23·14. Comment on the following statements—

(*i*) A wave-train of infinite extensions in time is a concept, but not a physical reality.

(*ii*) Fourier transform of a wave-group of length L involving wavelength λ_0 will have maximum $F(\lambda)$ at $\lambda=\lambda_0$ and spread $\Delta\lambda_{1/2}$ of the order of λ_0/L.

(*iii*) A slit of width a means transmission 1 over $-a/2<y<a/2$, and 0 outside this range. Fourier transform of this is the *sinc* function (§ 9·3)

23·15. Consider a 2 slit interference arrangement, in which a series of photo electric receivers are placed in the plane of obeervation. These are capable of recording the arrival of individual photons. Discuss what you may observe working with light of extremely low intensity. How are the results interpretted ? Suppose one places some device to find out from which of the two slits a given photon came out. How will be observations be affected ?

23·16. From eq. (23·20) write down those mathematical operators which, acting on wave function ψ, give us (*i*) the energy of the system, (*ii*) the linear momentum of the system.

23·17. If ψ is a normalized function, write down expressions for (*i*) probability of finding the particle is a certain volume range, and (*ii*) the expectation value of x component of angular momentum.

23·18. Examine Eq. (23·32). What would you conclude if R_{12} is found to be zero for $q=x$ and finite for $q=y$ (z being the direction of observation) ?

23·19. Compare the probability distribution of Fig. 23·9 with the expectations from classical theory.

23·20. Comment on the 'separation of variables' method used in solving the Schrodinger equation for *H*-atom. Could this be done in cartesian or any other coordinate system ? Upto Eq. (23·49) nothing is said about V, except that it is not a function of ϕ or θ. Does it mean that the solution thus far has wider validity?

24

The Special Theory of Relativity

If the *same event* is observed by two observers who are in *relative motion*, then their measurements would differ. In Galileo's idea if x', y', z', t' are the measuıements by observer S', who has velocity V relative to observer S, then the measurerements x, y, z, t as made by S (Fig. 24·1) are given by

$$\left.\begin{aligned} x&=x'+Vt' \\ y&=y' \\ z&=z' \\ t&=t' \end{aligned}\right\} \quad \text{or} \quad \left.\begin{aligned} x'&=x-Vt \\ y'&=y \\ z'&=z \\ t'&=t \end{aligned}\right\} \qquad \ldots(24{\cdot}1)$$

where axes x, y, z and x', y', z' are taken parallel, the motion of S' is along the x' axis, and the origin $0'$ crosses 0 at time $t'=t=0$. The set of Eq. (24·1) is called the *Galilean transformation* of coordinates between observers in uniform relative motion.

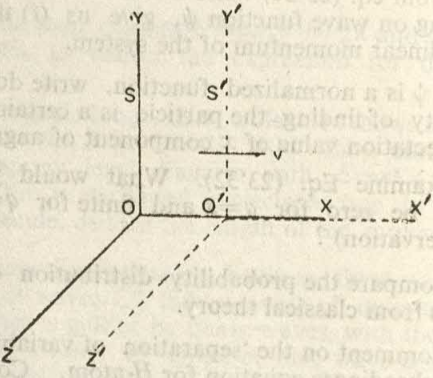


Fig. 24·1.
Two frames S, S′ in relative motion along x-axis

Newtonian mechanics accepted Galilean transformation, which leads to

$$\left.\begin{aligned} \Delta x&=\Delta x' \\ \Delta y&=\Delta y' \\ \Delta z&=\Delta z' \\ \Delta t&=\Delta t' \end{aligned}\right\} \qquad \ldots(24{\cdot}2)$$

and also to

$$\frac{dx}{dt}=\frac{dx'}{dt}+V=V'+V \qquad ...(24{\cdot}3)$$

Eqs. (24·2) give the Galilean transformation for intervals of space (lengths) and time and Eq. (24·3) is the law of addition of velocities.

Eq. (24·3) leads to

$$\frac{d^2x}{dt^2}=\frac{d^2x'}{dt'^2} \qquad ...(24{\cdot}4)$$

which is independent of V. The newtonian mechanics will apply identically[*] in the measurements made by S or S'. For this reason Eqs. (24·1) to (24·4) are said to represent *newtonian relativity*. Measurements x and dx/dt differ from x' and dx'/dt' respectively, but intervals Δx and $\triangle t$ and acceleration d^2x/dt^2 are identical with $\Delta x'$, $\Delta t'$, and d^2x/dt^2 repectively.

But careful experiments show that Galilean transformation is only an approximation ; it is not the complete truth. This chapter deals with the experimental evidence, and the theory which gave the new transformation : *Einstein's Special*[†] *Theory of Relativity*. As you will see newtonian relativity is good enough for $V<<c$ (speed of light) ; but when V/c is not insignificant we find $\triangle x'\neq\triangle x$, $\triangle t'\neq\triangle t$ and $dx'dt\neq dx'/dt'+V$. Philosophers were quite upset by such predictions. But in science the theory explained many discrepancies and predicted many new things. Among the most sepectacular predictions is the equation $\Delta E=c^2\Delta m$ regarding mass and energy interconversion.

24·1. Relativity and the Propagation of Light. If ether is accepted as the medium for the propagation of waves of light, then the waves have a speed *c relative to ether*. Now consider a telescope on earth viewing a star (Fig. 24·2). The telescope moves with earth's orbital speed V, which is about $10^{-4}c$. If we ignore V the image will be formed at P, the true focus. But if we take the telescope moving *towards* the star, then by the time the wave travels length L to P, we will find P at P', where $PP'=(V/c)PL$.

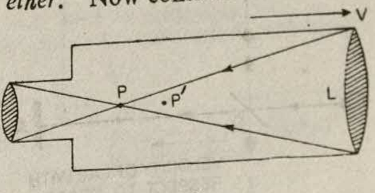


Fig. 24·2.
Effect of motion of a telescope through the ether

---

[*] If we also assume $m=m'$,

[†] The word 'special' is to distinguish it with the 'General' Theory of Relativity, which applied when the observers have non-zero accelerations also. The special theory deals with uniform relative motions only.

If the telescope is moving *away* from the star, the shift of P will be away from L by the same amount, so that by observations in two situations the difference should be observed. This effect was looked for by Arago long ago, but was not observed. Bradley looked for a similar angular deflection when V was perpendicular to the line of sight. Many such results, involving differences of the first order in V/c were expected, but were not observed.

Did it mean that each observer (like the telescope) carried ether along with itself ? Fresnel in 1818 produced arguments showing that a material medium moving with speed V dragged ether with itself to the extent of $V(1-n^{-2})$, where n is its refractive index. This is called the *ether drag*. Fizeau in 1851* showed by direct experiment that a column of water does drag the waves of light to this extent. In 1895 Lorentz deduced Fresnel's formula for ether drag, using the electromagnetic theory. Ether drag would mean that relative to the moving water, for example, ether would have speed Vn^{-2}. This will be the *ether wind*.

Michelson examined the possibility of using interference fringes to measure the *ether wind*. He found that any device used would only measure effects proportional to $(V/c)^2$, *i.e.*, the *second order* effect. Since earth's orbital speed is $\sim 10^{-4}c$, the effect in (say) 1 m length will amount to $\sim 10^{-8}\ m$, which is $\sim\lambda/50$ for visible light. Yet, he devised his famous interferometer (§ 12·6) for just this experiment. In 1887 he performed the crucial experiment along with Morley in which this effect, if present, could have been measured.

24·2. The Michelson and Morley Experiment. The Michelson interferometer has a significant feature that a beam of light

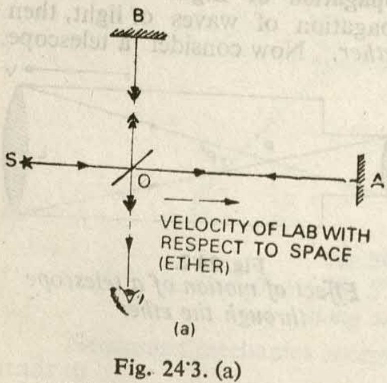


Fig. 24·3. (a)
Michelson interferometer with OA parallel to V

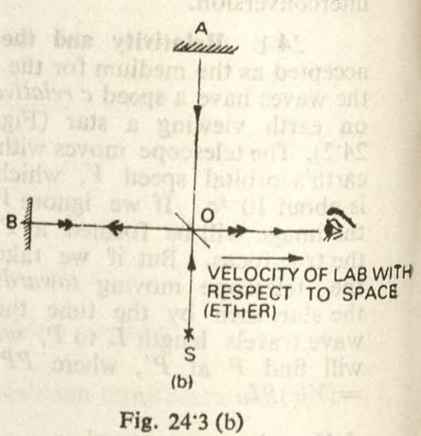


Fig. 24·3 (b)
Michelson interferometer with OB parallel to V

---

* Just 3 years after the first terrestrial method for measuring the speed of light was developed by Fixeau himself.

splits into two parts, which travel along *mutually perpendicular* directions, and then return to form interference fringes. Michelson and Morley mounted the interferometer on a rock floated on mercury, and keeping one arm OA (Fig. 24·3 *a*) parallel to earth's orbital velocity V, they got the interference fringes. Then the rock was turned by 90°, making arm OA perpendicular to V (Fig. 24·3 *b*), and a close watch was kept on the fringes. No shift of the fringes was observed.

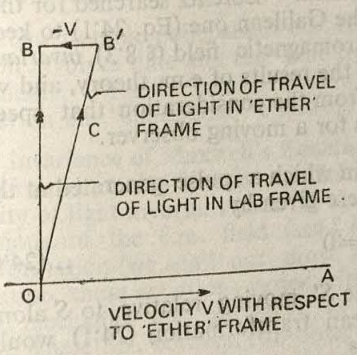
The theory, based on Galilean velocity addition, indicated that a change equal to path change $(OA+OB)V^2/c^2$ should have occurred. Making each distance $\sim 10m$, they expected a shift of n fringes, where with $(\lambda=5\times10^{-7}\text{ m})$

$$n=\frac{2\times10m}{5\times10^{-7}m}\times(10^{-4})^2\approx 0{\cdot}4$$

But Michelson and Morley state that the *observed shift* was "certainly less than one-twentieth part of the computed shift and possibly less than one-fortieth part". Quite reasonably, this was within the error of the observations. So it was a *null result*, *i.e.*, there is no ether wind at all.

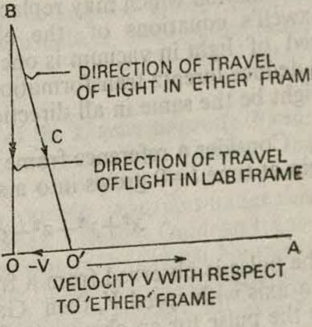
Theory of the Michelson-Morley Experiment. In Fig. 24·3 the pulse travelling from O to A approaches A at a speed $c-V$, and the return pulse from A to O approach O at a speed $c+V$, according to Galilean transformation. So the two-way time is

$$t_A=\frac{OA}{c-V}+\frac{OA}{c+V}=\frac{2OA}{c}\cdot\frac{1}{1-\beta^2}\text{. where }\beta=\frac{V}{c}.$$


-V
B
B′
DIRECTION OF TRAVEL OF LIGHT IN 'ETHER' FRAME
C
DIRECTION OF TRAVEL OF LIGHT IN LAB FRAME
A
O
VELOCITY V WITH RESPECT TO 'ETHER' FRAME

(a)

Fig. 24·4. (*a*)
Pulse from O reaching B′


B
DIRECTION OF TRAVEL OF LIGHT IN 'ETHER' FRAME
C
DIRECTION OF TRAVEL OF LIGHT IN LAB FRAME
O -V O′
A
VELOCITY V WITH RESPECT TO 'ETHER' FRAME

(b):

Fig. 24·4. (*b*)
Pulse from B reaching O′

For the pulse travelling towards B we note that by the time it travels there B will move to B', where magnitudewise $BB'/OB'=V/c$. Thus the pulse approaches B at speed $(c^2-V^2)^{1/2}$. The same speed applies for the return pules from B approaching O. So for this path the return time is

$$t_B=\frac{OB}{(c^2-V^2)^{1/2}}+\frac{OB}{(c^2-V^2)^{1/2}}=\frac{2\,OB}{c}\,\frac{1}{\sqrt{1-\beta^2}}$$

The time difference is thus given by

$$t_A-t_B=\frac{2\,OA}{c}\,\frac{1}{1-\beta^2}-\frac{2\,OB}{c}\,\frac{1}{\sqrt{1-\beta^2}}$$

Now comes the crucial step. Turning the two paths by 90° leads to the new time difference

$$t_A'-t_B'=\frac{2\,OA}{c}\,\frac{1}{\sqrt{1-\beta^2}}-\frac{2\,OB}{c}\,\frac{1}{1-\beta^2}$$

because OB is now parallel to V and OA is perpendicular to V. Thus the result of 90° turning is to change the time difference by

$$\Delta t\equiv(t_A'-t_B')-(t_A-t_B)=\frac{2(OA+OB)}{c}\left[\frac{1}{\sqrt{1-\beta^2}}-\frac{1}{1-\beta^2}\right]$$

$$=\frac{OA+OB}{c}\,\beta^2$$

using binomial expansion since $\beta<<1$. The equivalent change in path difference is

$$\Delta x=c\,\Delta t=(OA+OB)\,\beta^2 \qquad \text{...(24·5)}$$

The Michelson-Morley experiment and other variations of it have been repeated under many different circumstances, and the *null result* has been established fairly firmly. We therefore conclude that the concept of a fixed frame of reference (like 'ether' filling all space) cannot be checked by experiment and must therefore be discarded. Also, we conclude that *the speed of light in vacuum is the same in all frames of reference which are in uniform relative motion.*

24·3. The Lorentz Transformation. Lorentz searched for that transformation which may replace the Galilean one (Eq. 24·1) to keep Maxwell's equations of the electromagnetic field (§ 8·3) *invariant.* Speed of light in vacuum is one of the results of e.m. theory, and we will deduce lorentz transformation from the consideration that speed of light be the same in all directions for a moving observer.

Consider a reference frame S in which a pulse generated at the origin at time $t=0$ grows into a sphere given by

$$x^2+y^2+z^2-c^2t^2=0 \qquad \text{...(24·6)}$$

If the pulse is observed from a frame S', moving relative to S along the x-axis with speed V, then Galilean transformation (24·1) would give the pulse for an observe in S' as

$$(x'+Vt')^2+y'^2+z'^2-c^2t'^2=0$$

or

$$x'^2+2x'Vt'+V^2t'^2+y'^2+z'^2-c^2t'^2=0$$

This is not a sphere. The terms in y' and z' are of the proper form to give us a sphere, but the x' and t' terms are not. To get rid of the term $2x'\,Vt'$ let us change the time transformation from $t=t'$ to

$$t=t'+\frac{Vx'}{c^2} \qquad \text{...(24·7)}$$

Substitution of this in Eq. (24·6) now gives us

$$x'^2+2x'Vt'+V^2t'^2+y'^2+z'^2-c^2t'^2-2x'Vt'-\frac{V^2x'^2}{c^2}=0$$

$$\text{or}\quad x'^2\left(1-\frac{V^2}{c^2}\right)+y'^2+z'^2-c^2t'^2\left(1-\frac{V^2}{c^2}\right)=0$$

Now we find that a common factor $(1-V^2/c^2)$ occurs with both x'^2 and t'^2 terms. Hence the transformation for x as well as t needs a further factor $(1-V^2/c^2)^{1/2}$. Thus set of Eq. (24·1) is replaced by

$$\left.\begin{aligned} x&=\frac{x'+Vt'}{\sqrt{1-V^2/c^2}}=\gamma(x'+Vt') \\ y&=y' \\ z&=z' \\ t&=\frac{t'+\dfrac{Vx'}{c^2}}{\sqrt{1-V^2/c^2}}=\gamma\left(t'+\frac{Vx'}{c^2}\right) \end{aligned}\right\} \qquad \ldots(24{\cdot}8)$$

wher $\gamma=1/\sqrt{1-\beta^2}$. This is called the *Lorentz transformation for space and time coordinates* for frames in uniform relative motion. Substitution of these into Eq. (24·6) gives us

$$x'^2+y'^2+z'^2-c^2t'^2=0$$

which is a sphere, as desired. In other words, the pulse, as observed from frame S', also grows into a sphere at the rate c if the measurements are related by eq. (24·8) and not by eq. (24·1).

One notable point about eqs. (24.8) is that the space transformation involves time and the time transformation involves the space coordinate. Thus the transformation is not a "space *and* time" transformation but "space-time" transformation.

Invariance of Maxwell's Equations of the Electromagnetic Field. Transformation (24·8) has significance beyond just keeping the velocity of light invariant with the observer's motion. All Maxwell's equations of the e.m. field (see § 8·3) remain invariant under this transformation (we shall not prove it here). Under Galilean transformation these equations varied with the observer (though Newton's law of mechanics did not vary), *i.e.*, the laws of electrodynamics varied. Therefore, between two observers A and B in uniform relative motion, one could decide not only the relative velocity V_{AB}, but also the absolute velocities V_A and V_B with reference to some absolute reference frame. Lorentz transformation means that all frames in uniform relative motion show the same laws of electrodynamics, so that no experiment on electrodynamics can measure V_A and V_B ; only V_{AB} is measurable.

24·4. The Special Theory of Relativity. Einstein took the fundamental view that what cannot be measured by an experiment has no meaning in science. Hence the concept of an *absolute* reference frame in space is meaningless, and only *relative* motions are measurable and hence meaningful. Thus the concept of ether or any other 'fixed' frame of reference was discarded by him altogether.

The null result of the Michelson-Morley experiment on electromagnetic waves was extended by Einstein to include all conceivable experiments, including those in mechanics, and given the status of a fundamental postulate. In 1905 he put forth the following postulates of what is now called the Special Theory of Relativity :

1. *The laws of electrodynamics (including the propagation of light with a speed c in free space) and the laws of mechanics are exactly the same in all frames in uniform relative motion.*
2. *It is impossible to devise an experiment which may detect a state of absolute motion.*

The first postulate is equivalent to the null result of the Michelson-Morley experiment, but extends it to all conceivable experiments. The second says that not only the laws are identical, but also the physical constants like h, c, e, k would be the same for measurements in all frames in uniform relative motion.

Sometimes the postulates are put in three parts as follows ;

(*A*) For frames of reference in uniform relative motion, the speed of light is independent of the reference frame in which the measurement is made.

(*B*) The space is isotropic and uniform.

(*C*) The fundamental laws of physics are identical for any two observers in uniform relative motion.

In this form the null result of Michelson-Morley experiment in put separately in (*A*) and its extension to *all laws* of physics is stated separately in (*C*). Postulate (*B*) is really an additional statement included in postulate 2 given earlier but often not clearly appreciated. It is obvious that if space were not isotropic or not uniform, one could use the varying values of the constants as a device for determining absolute motion.

The Operational View of Relativity. *The theory of relativity is essentially a theory of measurement.* Consider, for instance, a railway train in uniform motion relative to the earth. Events occurring in the train are observed by observers in the train and also by observers on the earth ; similarly events occurring on the earth are observed by observers on the earth and also by observer in the train. So far we have not carefully examined *how* an observer in the train would measure the length of a rod on the ground, or how on observer on the ground would measure the time interval between two events occurring in the train ? The *operation* for such measurements has to be clearly specified. Einstein stated the operation as follows :

(*a*) The instruments in the two frames of reference cannot communicate directly except when they coincide in position.

(*b*) The only other communication can be through pulses of light.

If we proceed through this operational point of view we find that the transformation is not of the Galilean form, but of the Lorentz form. Thus the Lorentz transformation comes out as a consequence of the operational point of view about measurements of events in one reference frame from another frame in uniform relative motion. To emphasize this we shall use the operational procedure in one exemple (§ 24·6), although the equivalent results are more easily obtained from the Lorentz transformation.

24·5. Length Contraction and Time Dilation. An *event* is expressed by (x, y, z, t) or (x', y', z', t'). Customarily, the reference frame relative to which the event has no motion is denoted by the *primed* letter S'. The velocity of S' relative to reference frame S is denoted by V. In the railway-and-ground example, for a rod placed in the railway, S' is the railway-reference frame, S is the ground based frame ; for a rod lying on the ground, the ground-based frame is S' and railway-based frame is S. Uniform practice is important for the correct use of Eqs. (24·8).

Consider two positions, like the two ends of a rod, in frame S' whose coordinates *at a given time* t' as measured by an observer in the same frame S' are as $x_1' y_1' z_1'$ and $x_2' y_2' z_2'$. Then the intervals defined by

$$\triangle x' = x_2' - x_1' \text{ etc.}$$

are called the *proper space intervals.* If the same two position in S' frame are measured from frame S at the *same time* t (not t') the intervals defined by

$$\triangle x = x_2 - x_1 \text{ etc.}$$

are called the *non-proper* space *intervals.* Thus the length of a stick placed in a moving railway as measured by a person in the railway itself is a proper length, while the length of the same stick measured by an observer standing on the earth is non-proper .

Similarly, if two events occur at the *same place* in frame S′, the time interval

$$\triangle t' = t_2' - t_1'$$

is called the *proper time interval.* If the same pair of events in S' frame are observed fram a frame S and are found to occur at the same place x', y', z' (not x, y, z) at times t_1 and t_2 , then the interval

$$\Delta t = t_2 - t_1$$

is called the corresponding *non-proper time interval.* Let us proceed to establish the relationship between the proper intervals $\triangle x'$, $\triangle y'$ $\Delta z'$, $\triangle t'$ and the corresponding non-proper intervals $\angle x$, Δy, Δz, $\triangle t$.

For space intervals, the proper intervals are obtained by writing the first three Lorentz equations (24·8) *for same t.* The last equation in (24·8) gives

$$t' = \frac{t}{\gamma} - \frac{Vx'}{c^2}$$

so that the equation for x in (24·8) becomes

$$x=\gamma\left[x'+\frac{Vt}{\gamma}-\frac{V^2x'}{c^2}\right]=\gamma\left[\frac{x'}{\gamma^2}+\frac{Vt}{\gamma}\right]$$

or $$x=\frac{x'}{\gamma}+Vt \qquad \text{...(24·9)}$$

This expresses x in terms of x' and t (not t'). Writting out for x_1 and x_2 and taking the difference we at once get

$$\triangle x=x_2-x_1=\frac{x'_2-x_2'}{\gamma}=\frac{\triangle x'}{y}$$

or $$\boxed{\triangle x=\triangle x'\sqrt{1-V^2/c^2}} \qquad \text{...(24·10)}$$

Thus the non-proper length $\triangle x$ in the direction of relative motion is *shorter* than the proper length $\triangle x'$ by a factor $\sqrt{1-V^2/c^2}$. This is called *length contraction* in special relativity.

The lengths in directions perpendicular to that of relative motion are equal in both frames. That is, $\triangle y=\triangle y'$ and $\triangle z=\triangle z'$.

For time intervals the last equation in (24.8) gives, *for the same position* x',

$$t_1=\gamma\left(t_1'+\frac{Vx'}{c^2}\right), \qquad t_2=\gamma\left(t_2'+\frac{Vx'}{c^2}\right)$$

whence we get

$$\triangle t\equiv t_2-t_1=\gamma\,(t_2'-t_1')=\gamma\,\triangle t'$$

or $$\boxed{\triangle t=\frac{\triangle t'}{\sqrt{1-V^2/c^2}}} \qquad \text{...(24·11)}$$

Thus the non-proper time interval $\triangle t$ is *larger* than the proper time interval by a factor $1/\sqrt{1-V^2/c^2}$. This is called *time dilation* in special relativity.

Notice that for non-proper measurement of length the measurements are made at the same time t. Hence we needed the conversion from the first equation of (24·8) to eq. (24·9). In the non-proper measurement of time the measurements t_1 and t_2 are for same position x', y' z' (not x, y, z), hence the last equation in (24·8) could be directly used.

Example 24·1. A reference frame S' moves along the x-axis with velocity 300 m/sec relative to a frame S', their origins coinciding at $t=t'=0$. If an event occurring at $x'=10$ m, $y'=5$ m. $z'=5$ m, $t'=5$ sec, is observed from frame S deduce x, y, z, t.

Solution. The data give

$$Vt'=1500\text{ m}\ ;\ Vx'=3000\text{ m}^2\text{s}^{-1}\ ;\ \gamma=\left[1-\left(\frac{300}{3\times10^8}\right)^2\right]^{-1/2}=1+\frac{10^{-12}}{2}$$

Substitution in Eq. (24·8) yields

$$x=(x'+Vt')\gamma=1510\gamma\ m\ ;\ y=y'=5\ m;\ z=z'=2\ m$$

$$t=\left(t'+\frac{Vx'}{c^2}\right)\gamma=\left(5+\frac{10^{-13}}{3}\right)\gamma s$$

[Thus x and x' differ by large amounts, y and z do not differ at all from y' and z' respectively, while t and t' differ but by negligible amounts.] ⊙

Example 24·2. Repeat Ex. 24·1 with $x'=10^{18}$ m. rest of the data remaining the same. (Star distances are of this ordder).

Solution. $x=(x'+Vt')\,\gamma=(10^{18}+1500)\,\gamma\ m\ ;\ \gamma=1+\frac{10^{-12}}{2}$

$$t=t'+Vx'/c^2=(5+333)\gamma\ s$$

[In this case due to large x' we find t to be several times t' even though γ is still quite close to unity].

Example 24·3. A rod placed along the length in a railway train measures 2 metres for an observer standing on the ground. How much would it measure for an observer in the train itself? Speed of the train in 30 m/sec.

Solution. The given length is the non-proper length to be designated by Δx. To deduce the proper length $\Delta x'$ we use eq. (24·10)

$$\Delta x'=\frac{\Delta x}{\sqrt{1-V^2/c^2}}=\frac{2}{\sqrt{1-10^{-14}}}=(2+\tfrac{1}{2}\times10^{-14})\text{ metre}$$

[The change is negligable, but the point is that $\Delta x'$ is *greater* than Δx].

Example 24·4. A laboratory observation shows that a length 2 metre is crossed by a beam of radioactive particles in time $1{\cdot}0\times10^{-8}$ second and in this process half the particles disintegrate. Deduce the proper half-life of the particles.

Solution. The event (of disintegration) occurs *in the particles*, whose speed in the laboratory is, from the given data. $V=2\times10^8\ m/s$.

The half-life as measured *in the lab* is $1{\cdot}0\times10^{-8}$ second. This is the non-proper measurement, to be designated Δt. The proper half-life $\Delta t'$ is therefore given by eq. (24·11):

$$\Delta t=\Delta t\sqrt{1-V^2/c^2}=10^{-8}\sqrt{1-(2/3)^2}$$

$$=0{\cdot}78\times10^{-8}\ \text{s}.$$

Example 24·5. In Example 24·4 deduce the traversed length as seen by an observer fixed relative to the particle. Use it to deduce the half-life.

Solution. The path in the lab as measured by a lab observer is 2 metres. So this is tthe proper length $\Delta x'$. The length of this as seen by the moving observer would be the non-proper length Δx. Hence, using eq. (24·10).

$$\Delta x=200\sqrt{1-(2/3)^2}=1{\cdot}56\ \text{m}$$

The half-life, as measured by an observer in the particle frame, is $\Delta x/V$, and this is the proper half-life, to be designated by $\Delta t'$: Hence

$$\Delta t'=\frac{1{\cdot}56\ m}{2\times10^8\ ms^{-1}}=0{\cdot}78\times10^{-8}.$$

24·6. The Operational View of Time Dilation. Fig. 24·5 shows a reference frame S' moving relative to another frame S with speed V along x' axis. A pair of events at the same place in S' occurs at time interval $\Delta t'$, and we have to see the *operation* how an observer in frame S will measure this time interval. Recall rules (*a*) and (*b*) laid down by Einstein (§ 24·4) that a communication to S can occur only when the instruments coincide in position, or through pulses of light.

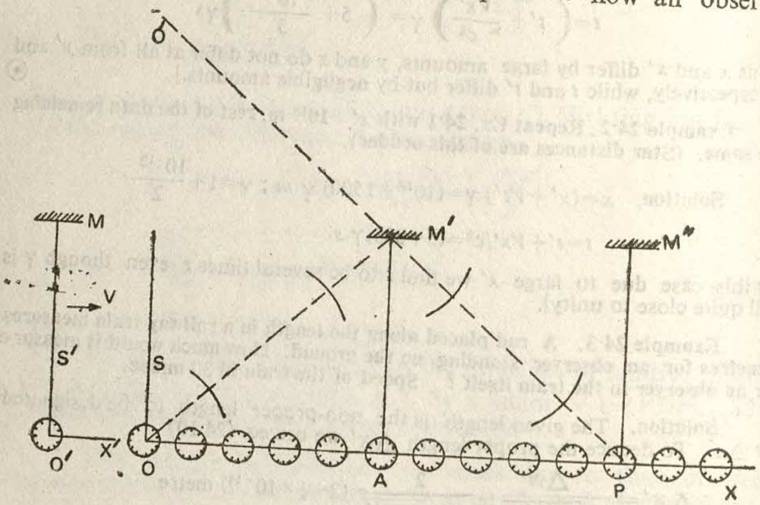


Fig. 24·5.

The Operational View of Time Dilation

So we consider a pair of events involving pulses of light. Frame S' contains a clock at O' and a mirror M at transverse distance y. Frame S contains a chain of clocks along the x-axis ; these will start when a pulse of light reaches there, and stop when the next pulse reaches there. The pulse of light is created just as O' coincides with O during the motion of S'.

Now, for frame S', the mirror M is all the time along perpendicular to x' axis at O'. The pulse starting from O' and returning to it after reflection from M takes time

$$\Delta t' = \frac{2y}{c} \qquad \text{...(24·12)}$$

For frame S the pulse starts from O, and as it grows on, the mirror M (fixed in S' frame) moves along the x-axis at speed V. So the pulse contacts M at a position like M' and the reflected pulse reaches a particular clock P to stop it. (The S' frame has its MO' at position $M''P$ just at this instant.)

What interval Δt does the observer in S frame record ? The event started at O and ended at P in this frame. The 'start' signal came to S frame by method (*a*), *i.e.*, clock O' coinciding with O. The 'stop' signal came to S frame by method (*b*), *i.e.*. the pulse of light. Clock P also is started by the direct pulse, which reached there at time OP/c later than the start of the event. We presume that this delay is taken care of by synchronising the clocks in the chain appropriately. With this understanding the time $\triangle t$ is measured by clock P in frame S.

By Einstein's postulate of constancy of the speed of light, $OM'+MP=c\triangle t$. Simple geometry gives now the relation between $\triangle t$ and $\triangle t'$.

$$AM'^2=OM'^2-OA^2=(\tfrac{1}{2}c\triangle t)^2-(\tfrac{1}{3}V\triangle t)^2=(\tfrac{1}{2}c\triangle t)^2\left(1-\frac{V^2}{c^2}\right)$$

But Eq. (24·12) gives $\frac{1}{2}c\triangle t'=y=AM'$. Hence

$$(\tfrac{1}{2}c\triangle t')^2=(\tfrac{1}{2}c\triangle t)^2\left(1-\frac{V^2}{c^2}\right)$$

$$\therefore \qquad \triangle t=\frac{\triangle t'}{\sqrt{1-V^2/c^2}}$$

This is the same as Eq. (24·11), deduced from Lorentz equations.

24·7. Experimental Verification of Time Dilation. Time dilation is best verified in experiments on decay half-life of fast moving nuclear particles. π^+ meson is one such which can travel in the laboratory at quite high speeds, close to c, and which decays with time.

Now, we measure the time of travel $\triangle t$ of a π^+ meson beam across length L in the laboratory. We also determine the π^+ flux at the start and end of this length, and from this we get the half-life $\tau_{1/2}$ for their decay. Observations show that the $\tau_{1/2}$ thus measured varies with velocity V $(=L/\triangle t)$ of the particles. The larger the V value, the greater is the measured $\tau_{1/2}$. Nuclear decay time should be a characteristic constant. So how do we understand this varying $\tau'_{1/2}$? Eq. (24·11) gives the answer. Convert $\tau_{1/2}$ (measured) to $\tau_{1/2}$ (proper half-life) using.

$$\tau'_{1/2}=\tau_{1/2}\sqrt{1-V^2/c^2}$$

Then you find that for all the varying $\tau_{1/2}$ for π^+ mesons the $\tau i_{1/2}$ value comes out the same $1·8\times10^{-8}$ *s*.

As one example, a π^+ meson beam produced with speed 0·990c travelling a 30 m distance decreased to 58% of the original flux. Using

$$2^{-\triangle t/\tau_{1/2}}=0·58$$

gives $\tau_{1/2}=12·8\times10^{-8}$ *s*. But $\tau'_{1/2}$ comes to $1·8\times10^{-8}$ *s*.

24·8. The Relativistic Transformation of Velocities. Let v_x', v_y', v_z' be the components of velocities of a particle as measured in a frame S'. If frame S' has a uniform speed V along the x-axis of a frame S, we have to deduce the velocity components v_x, v_y, v_z of that particle as measured in frame S. It is assumed that the x, y, z axes are parallel to x', y', z' axes, respectively.

Since v_x is defined as dx/dt we have to get dx and dt from the Lorentz transformation* (24·8) in terms of dx' and dt'.

$$dx=\gamma dx'+\gamma V dt' \;; \qquad dt=\gamma dt'+\frac{V}{c^2}\,\gamma dx'$$

$$\therefore \quad v_x=\frac{dx'+Vdt'}{dt'+\frac{V}{c^2}\,dx'}=\left(v_x'+V\right)\left[1+\frac{v_x'+V}{c^2}\right]^{-1} \qquad \text{...(24·13}a)$$

Similarly, we may deduce

$$v_y=v_y'\left[\gamma\left(1+\frac{v_x'V}{c^2}\right)\right]^{-1} ; \quad v_z=v'_z\left[\gamma\left(1+\frac{v_x'V}{c^2}\right)\right]^{-1} \qquad \text{...(24·13}b)$$

Derivation of the inverse transformation is left to the reader. Note that v_y, v_z also are affected by V, unlike x and y.

Example 24.6. In the laboratory one particle *A* moves with speed $+0.9c$ along the x-axis, while another particle *B* moves with speed—0·8*c* along the x-axis. Deduce the velocity of *A* relative to *B*.

Solution. Consider a frame moving with particle *B*, Relative to this B-frame the laboratory frame moves with speed 0·8*c* Then we have

$$v_x'=0·9c \text{ and } V=0·8c.$$

Substitution in eq. (24·13 *a*) gives

$$v_0=\frac{(0·9+0·8)c}{1+\frac{0·9\times 0·8}{1}}=\frac{1·7}{1·72}\,c=0·988c.$$

A special feature of Eq. (24·13a) is that if $v'_x=c$ comes than v_x out to be just *c*, whatever the value of *V* may be. This, in fact, was the result of Michelson-Morley experiment. In effect this equation means that *in no reference framecan the observed velocity exeed c*; the velocity of light in vacuum is nature's upper limit to velocities. At the cther extreme, let $v_x'=c$ and $V=-(c-\delta)$, with $\delta \to 0$; one gets $v_x=c$. Thus substracting *c* from *c* leads to *c* itself. In effect it only means the V is not measurable even when it approaches *c* and is opposite to *c* in direction.

24·9. Variation of Mass with Velocity. In newtonian mechanics linear momentum *p* is defined by $p=m_o V$, where m_o is mass of the body and is treated constant for a given body, irrespective of its speed.

Now conservation of momentum is a fundamental law of mechanics and according to the postulates of Einstein's relativity it must hold in Lorentz transformation also. For this purpose the *definition* of momentum has to be revised to $p=mV$, with

---

* One may be tempted to use Eqs. (24·10) and (24·11). But these refer to two x′ measurements at the same time, and two *t*′ measurements at the same place, respectively. The velocity problem refers to change in x′ with *t*′.

$$\boxed{m \equiv \frac{m_o}{(1-V^2/c^2)^{1/2}} = \gamma m_o} \qquad \text{...(24·14)}$$

m_o is called the '*rest mass*' or proper mass of the body, while m is called the *relativistic mass*, which is the effective value of mass in a frame of reference relative to which the body has speed V.

Proof of eq. (24·14) is as follows. Consider the collision of two particles A, B of *equal masses*, which is symmetrical in a reference frame S', as shown in fig. 24·6*a*. By 'symmetrical' we mean that if the initial velocity components are v_x', v_y' for A, they are $-v_x'$, $-v_y'$ for B. After the collision the velocity components along x-axis do not change, while those for y'-axis reverse in sign for *both* A and B.

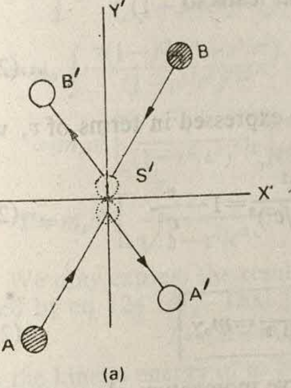


Fig. 24·6 (*a*)
Collison of A and B in frame S′, where it is symmetrical

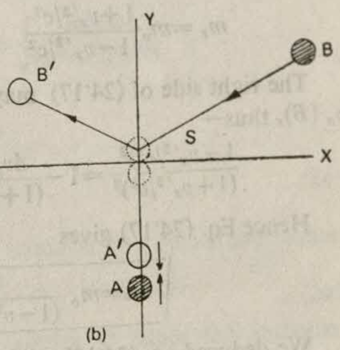


Fig. 24·6 (*b*)
Collision of fig. 24·6 (a) in frame S where $(v_x)_A$ is zero

If the rest masses of A and B are m_o *each*, then their masses m_A and m_B under velocities v_x' and $-v_x'$ respectively could be different from one another. But symmetry considerations require that Σp_x be neither positive nor negative, and that settles $m_A = m_B$, though they need not be equal to m_o. Dependence of m on $|v_x'|$ is yet to be obtained.

Now consider the same collision in a reference frame S moving with velocity v_x' relative to S'. This means $V = -v_x'$ and leads to x-components of velocities as follows :

$$\left.\begin{aligned} &\textit{Initial} \quad v_x\,(A)=0 \qquad v_x\,(B) = -\frac{2v_x'}{1+v_x'^2/c^2} \\ &\textit{Final} \quad v_x(A)=0 \qquad v_x\,(B) = -\frac{2v_x'}{1+v_x'^2/c^2} \end{aligned}\right\} \qquad \text{...(24·15)}$$

In this reference frame $m_A = m_o$ because there is no x-motion for particle A. But $m_B \neq m_o$ because particle B has a velocity $|v| = 2v_x'/(1+v_x'^2/c^2$. But conservation of p_x is unable to give the dependence of m_B on v.

So we proceed to y-components of velocities in the S frame. We get from (24·13 b),

$$\left.\begin{array}{ll} \textit{Initial } v_y(\text{A})=\dfrac{1}{\gamma}\,\dfrac{v_y'}{1-v_x'^2/c^2} & v_v\,(\text{B})=\dfrac{1}{\gamma}\,\dfrac{-v_y'}{1+v'_x{}^2/c^2} \\ \textit{Final } v_y\,(\text{A})=\dfrac{1}{\gamma}\,\dfrac{v_y'}{1-v_x'^2/c^2} & v_y\,(\text{B})=\dfrac{1}{\gamma}\,\dfrac{v_y'}{1+v_x'^2/c^2} \end{array}\right\} \quad \ldots(24{\cdot}16)$$

Conservation of p_y leads to

$$m_o\frac{1}{\gamma}\,\frac{-2v_y'}{1-v_x'^2/c^2}+m_v\frac{1}{\gamma}\,\frac{2v_y'}{1+v_x'^2/c^2}=0$$

where we have put $m_A=m_o$ and $m_B=m_V$ in view of the velocities of A and B respectively in frame S. That leads to

$$m_v=m_o\frac{1+v_x'^2/c^2}{1-v_x'^2/c^2} \quad \ldots(24\text{-}17)$$

The right side of (24·17) may be expressed in terms of v, which is v_x (B), thus—

$$\frac{1-v_x'^2/c^2)^2}{(1+v_x'^2/c^2)^2}=1-\frac{4v_x'^2/c^2}{(1+v_x'^2/c^2)^2}=1-\frac{v^2}{c^2} \quad \ldots(24{\cdot}18)$$

Hence Eq. (24·17) gives

$$\boxed{m_v=m_o\,\frac{1}{(1-v^2/c^2)^{1/2}}=m_o\gamma} \quad \ldots(24{\cdot}19)$$

We deduced eq. (24·19) to conserve momentum along the y-axis. But with the same changes of mass the x-component of momentum would still be conserved. Hence it is reasonable to introduce eq. (24·19) in a general way for all uniform motion.

Example 24·7. Deduce the velocity at which the mass of a particle becomes 10 times its rest mass.

Solution.

$$\frac{m}{m_0}=\frac{1}{(1-V^2/c^2)^{1/2}}=10$$

$$\therefore \quad 1-V^2/c^2=0{\cdot}01$$

$$\text{or} \quad V/c=(1-{\cdot}01)^{1/2}=(1-{\cdot}005)$$

$$\therefore \quad V=0{\cdot}995c.$$

24·10. Relativistic Force and Kinetic Energy. Force is defined by the *rate of change of momentum* of the body on which it acts. Also kinetic energy is defined as the integral of force times distance from state $v=0$ to the given v. These definitions are retained in relativistic mechanics, and they do not violate the postulates of relativity. For linear motion

$$F\equiv\frac{d}{dt}\,(mv)=m\frac{dv}{dt}+v\frac{dm}{dt} \quad \ldots(24{\cdot}20)$$

The appearance of the second term $v\,dm/dt$ is to be noted. For this reason in relativistic mechanics we often prefer to attack problems in terms of momentum rather than acceleration.

For kinetic energy T we write

$$T=\int_{o}^{v} F\,ds=\int\frac{d}{dt}(mv).\frac{ds}{dt}\,dt=\int\frac{ds}{dt}.\,d(mv)$$

Now ds/dt=the instantaneous velocity v. Also m is given by eq. (24·19). Hence

$$T=\int v.\,d\left[\frac{m_o v}{(1-v^2/c^2)^{1/2}}\right]=m_o\int\left[\frac{v}{(1-v^2/c^2)^{1/2}}+\frac{v^8/c^2}{(1-v^2/c^2)^{3/2}}\right]dv$$

$$=m_o\int\frac{v(1-v^2/c^2+v^2/c^2)}{(1-v^2/c^2)^{3/2}}\,dv=\int\frac{m_o v\,dv}{(1-v^2/c^2)^{3/2}}$$

$$=m_o c^2\left[\frac{1}{(1-v^2/c^2)^{1/2}}\right]_{v=o}^{v=v}$$

or $$T=m_o c^2\left[\frac{1}{\sqrt{1-v^2/c^2}}-1\right] \qquad \text{...(24·21)}$$

We may express the result in terms of the relativistic mass m defined by eq. (24·14). That gives

$$T=(m-m_o)\,c^2 \qquad \text{...(24·22)}$$

Thus the kinetic energy of a body equals c^2 times the increase of mass.

Example 24·8. Deduce the speed of an electron accelerated to a potential of 1 million volts.

Solution. Non-relativistic equation $T=\frac{1}{2}mv^2$ cannot be used here. Eq. (24·21), when rearranged, gives

$$\frac{1}{1-V^2/c^2}=\left(1+\frac{T}{m_0c^2}\right)^2 \qquad \text{...(24·23)}$$

We have

$$m_0c^2=9{\cdot}1\times10^{-31}\times(3{\cdot}0\times10^8)^2\ \text{joule}=\frac{8{\cdot}2\times10^{-14}}{1{\cdot}6\times10^{-19}}\,ev=0{\cdot}51\ Mev^*$$

With T=1·00 *Mev* (given), we have

$$\frac{1}{1-V^2/c^2}=\left(1+\frac{1{\cdot}00}{0{\cdot}51}\right)^2\Rightarrow V=0{\cdot}94c$$

Example 24·9. Calculate the relative values of kinetic energies of a particle accelerated to a seed of 0·500c, 0·900c, and 0·980c,

Solution.

$$T=m_0c^2\left[\frac{1}{(1-V^2/c^2)^{1/2}}-1\right]$$

* This is called *rest mass* of the electron, expressed in *Mev*.

Since m_0c^2 is common for all three cases, the bracketted term alone determines the relative values of kinetic energies. The values of this term give

$$T_1 : T_2 : T_3 = 0{\cdot}16 : 1{\cdot}27 : 4{\cdot}0 \text{ (check it)}$$

[Note that while the speed increased from $0{\cdot}500c$ to $0{\cdot}980c$ which is less than 2-fold, the kinetic energy increased 25-fold).

Example 24·10. The temperature of a body of specific heat 0·2 is raised by 200°C. Calculate the fractional increase of its mass.

Solution.

$$\triangle E = ms\triangle\theta = m\times(0{\cdot}2\times10^3\times4{\cdot}2)\times200 = 1{\cdot}7\times10^5\ m \text{ joule}$$

$$\frac{\triangle m}{m} = \frac{\triangle E}{mc^2} = \frac{1{\cdot}7\times10^5 m}{9\times10^{16}\ m} \approx 2\times10^{-12}$$

Thus mass of a 1 tonne body will increase by 2 microgram, which is too small for all practical purposes.

Example 24·11. A nucleus of mass 435·124 a. m. u. breaks into iwo nuclei of masses 97·936 and 135·951 a. m. u. and a neutron of mass 1·009 a. m. u. Calculate the energy released when 1 milligram of such nuclei distintegrate.

Solution. Loss of mass per fission$=0{\cdot}228$ a. m. u. (check)

$$\therefore \quad \text{fractional loss of mass} = \frac{{\cdot}228}{235} = 1{\cdot}0\times10^{-3}$$

$\therefore$ loss of mass in 1 mg$=1{\cdot}0\times10^{-9}$ kg

Energy released$=\triangle mc^2 = 1{\cdot}0\times10^{-9}\times9\times10^{16} = 9\times10^7$ joule ⊙

24·11. Equivalence of mass and Energy. In Eq. (24·22) we may consider m_oc^2 as an energy associated with mass m_o. This is called the '*rest energy*', and is in addition to all other earlier known forms of energy. The sum of rest energy and kinetic energy then becomes

$$E = mc^2 \qquad \ldots(24{\cdot}24)$$

An alternative way of showing that such a hypothesis is a *necessity* is as follows.

Consider a '*massless box*' containing a number of particles of different masses and momenta. A solid body fits this description if the geometrical form alone is taken as the massless box and the atoms and molecules in it are taken as the contained particles with various masses and momenta. In a reference frame fixed with the 'box' the total momentum of the particles is zero. If the i^{th} particle has mass m_i and kinetic energy T_i we would say

$$\text{total mass } m_o = \Sigma m_i$$

$$\text{total kinetic energy } T = \Sigma T_i$$

where the summation extends over all the particles.

In the *same frame of reference*, we can have an alternative description of the same system. We may consider the 'box with the particles' (the solid as such in our example) as a collective whole.

Now in this description the system has *no kenetic energy.* Yet, if we take the total mass as just Σm_i, that is m_o, there will be discrepancy about energy. Hence to keep the law of conservation of energy intact we must take the mass m, in the second description, such that

$$m-m_o=\frac{T}{c^2}$$

or $$c^2(m-m_o)=T=\Sigma T_i \qquad \text{...(24·24}a\text{)}$$

The right hand side expresses the sum of kinetic energies of the separate particles in the first description. When we switch over to the collective description, this appears in the form of increase of mass $\triangle m=m-m_o=T/c^2$.

However, energy can occur in various forms, the equivalence of which has been well established. Hence the statement may be generalised. We write Eq. (24·24) now as

$$E=mc^2=m_oc^2+T+E_1+E_2+\ldots\ldots$$

where E_1, E_2, etc. are energies in other forms like electrical energy, gravitational energy, etc., and E is the 'total energy' including m_oc^2. From this we put most generally

$$c^2\triangle m=\triangle T+\triangle E_1+\triangle E_2+\ldots \qquad (24{\cdot}25)$$

which means that a change of mass $\triangle m$ may appear in the form of energies of various kinds, whose total is $c^2\triangle m$.

The most spectacular examples of mass-energy equivalence are the *pair production* of electron and positron from γ-rays and the *annihilation* of a particle and its anti-particle to produce a γ-ray photon. In the first case new material particles are generated from energy of an electromagnetic field, while in the second case a pair of material particles is totally destroyed and equivalent energy appears as a photon.

An electron and its anti-particle (the positron) each has mass 9×10^{-31} kg whose mc^2 value converted to electron-volts is 0·51 Mev. Hence the minimum energy of a γ-ray photon to produce an electron-positron pair is $2\times0{\cdot}51=1{\cdot}02$ Mev.

PROBLEMS

24·1. If a reference frame A moves relative to another frame B with velocity $V_x\mathbf{i}+V_y\mathbf{j}+V_z\mathbf{k}$, write down the Lorentz transformation for x_a, y_a, z_a, t_a in terms of x_b, y_b, z_b, t_b and the velocity of frame A. The axes in the two frames are orthogonal and parallel, and the origins coincide at time $t_a=t_b=0$.

24·2. Show that for $V/c\ll1$, the Lorentz transformation reduces to the Newtonian (or Galilean) one.

24·3. If $\beta = V/c$ and $\gamma = 1/\sqrt{1-\beta^8}$, obtain a plot of γ against β. Also, if we define $\alpha = 1-\beta$ express γ in terms of α for $\alpha \ll 1$. Deduce the values of α, β and V when γ takes a value 10.

24·4 Consider the four-dimensional volume element $\Delta x\, \Delta y\, \Delta z \Delta t$. Show that this remains invariant under Galilean transformation as well as Lorentz transformation.

24·5 Maxwell's equations of the electromagnetic field are given in Chapter 8. Show that these are invariant in the Lorentz transformation but not in the Galilean one.

24·6 A rod travels parallel to its length with a speed $2·5 \times 10^7$ m/sec in a frame *S*. If the length as measured by a person in frame *S* is 25·0 cm, deduce the length as measured by a person riding on the rod.

24·7. Define *proper and non-proper* intervals of space and time, using the example of events on the ground and a jet plane in uniform motion. A carrom board measuring 1·00 m × 1·00 m when at rest on the ground is placed in the jet plane. How would it appear to players in the plane when the plane has velocity 1 km/s ?

24·8 In the theory of the Michelson-Morley experiment one does not need *OA* and *OB* to be equal. Examine why the difference of *OA* and *OB* must not be large. [Hint : See Chapter 21].

24·9. Examine the postulates *A*, *B*, *C* of §24·4. What is the meaning of postulate *B*, and why is it needed ? Comment on the statement that laws of physics are not the same in different frames of reference.

24·10. Prove that any velocity added to *c* gives *c* itself. Prove also that any velocity subtracted from *c* also gives *c* itself.

24·11. In a laboratory one particle *A* travels with velocity $+2·5 \times 10^8$ m/sec, another particle *B* travels with velocity $-2·9 \times 10^8$ m/sec, both along *x*-axis. Deduce the velocity of *A* relativity to *B*.

24·12. A square frame 1·0 m × 1·0m travels parallel to one of its sides with speed $2·8 \times 10^8$ m/sec. Deduce how the frame would appear to a lab observer.

24·13. The half-life of a radioactive particle is $2·0 \times 10^8$ sec. Deduce the fraction of such particles which can survive after travelling 2·0 m distance in the laboratory when their speed is (*i*) $2·0 \times 10^5$ m/sec, (*ii*) $2·0 \times 10^8$ m/sec, (*iii*) $2·96 \times 10^8$ m/sec.

24·14. As a beam of particles crosses 12·0 m length of a lab in $4·14 \times 10^{-8}$ *s*, half of these particles decay. Deduce the proper half-life of the particles. What would an agency moving with the particle get as a measure for the length of the lab ?

24·15. Without relativistic transformations the measurements of e/m of fundamental particles and life-times of decaying particles would be in a mess. Explain.

24·16. Deduce that kinetic energy T is given by $(m-m_o)c^2$, and that $E^2=m_o^2c^4+p^2c^2$.

24·17. Compute the speed of an electron when it is accelerated to (*i*) 10^5 volt, (*ii*) $2{\cdot}0\times10^6$ volt, (*iii*) $8{\cdot}0\times10^6$ volt.

24·18. Deduce the momentum of a proton with kinetic energy 2×10^3 Mev. (proton rest mass$=1{\cdot}673\times10^{-27}$ kg).

24·19. Show that at relativistic speed v, the force vs acceleration relation becomes $F=m_oa/(1-v^2/c^2)^{3/2}$. [Hint : See Eq. 24·20].

24·10. Prove that total energy E and momentum p are related by $E^2=(pc)^2+(m_oc^2)^2$. Show that for $E \gg m_oc^2$ this reduces to $p=E/c$, and for kinetic energy $K \ll m_oc^2$ it leads to $K=p^2/2m_o$.

25

Visual Photometry

'Photometry' in its general meaning involves the measurement of radiant energy. However, out of the total electromagnetic spectrum the human eye is sensitive to only a small region of wavelengths and even within this region the visual sensation for equal radiant energy is not uniform. For this reason photometry in terms of *visual sensation* is very different from that based on measurement of radiant energy. In this Chapter we will be dealing with visual photometry only. The final measuring device is then the human eye. Therefore, we will also include a brief description of the eye and its functioning.

25·1. Quantities Related with Visual Photometry

Table 25·1 gives a list of the various quantities related with visual sensation. The basic measure of *visual sensation* created by the total radiation emanating from a source or incident on a receiver is called the luminous flux, expressed by the symbol ϕ, and measured in a unit *lumen* (*lm*). We talk of a given area S of a source giving out luminous flux ϕ *lm* in a specified solid angle Ω. We also talk of a screen of area S receiving luminous flux ϕ *lm* from one or many sources.

TABLE 25·1

Quantities in Visual Photometry

Quantity	Definition	Symbol	Unit	Dimensions*
Luminous flux	A measure of the *total visual sensation* produced by given radiation in any solid angle Ω or surface S	ϕ	Lumen (*lm*)	[*lm*]
Luminous Intensity	Luminous flux emitted per unit solid angle : $d\phi/d\Omega$ (Applicable to a 'point source')	I_θ	Candela (cd)	$[lm\, L_x^2 L_y^{-2}]$

* Although the fundamental unit is defined for I (candela), we express the dimensions in terms of lumen (*lm*), which is more directly related to visual sensation.

Brightness	Luminous flux emitted per unit solid angle per unit area of the source transverse to the line of sight : dI/dS', where $dS'=dS \cos\theta$. (Applicable to a source of finite area)	B_θ	cd/m^2	$[lm\ L_x^2 L_y^{-4}]$
Luminance	Total luminous flux emitted per unit area of the source : $d\phi/dS$.	L	lm/m^2	$[lm\ L_y^{-2}]$
Illuminance	Total luminous flux received per unit area of a screen : $d\phi/dS$	E	lm/m^2 (=*lux*, *lx*)	$[lm\ L_y^{-2}]$

Out of the quantities defined, brightness (B) and intensity (I) are direction dependent. In contrast, Luminance is the integrated effect of B summed up for all directions. Also, for a source of finite dimensions, I is the integrated effect of B over the entire area of the source. Finally ϕ is the integrated effect of I over any given solid angle or any receiving area.

It may be noted that if solid angle is taken to be dimensionless, then there can be considerable mix-up in units. For this reason, we take the dimensions of solid angle as ($L_y^2 L_x^{-2}$), where L_x and L_y are lengths along and transverse to the ray, respectively.

Candela, the Fundamental SI Unit of Intensity

A 'standard candle' was used early to define intensity. But the latest fundamental unit is *Candela* defined as the *luminous flux emitted per unit solid angle along normal to the surface by one-sixtieth cm² of a blackbody radiator kept at the temperature of solidification of platinum* (2,046 K).

Thus 1 cm² of this blackbody surface at 2,046 K will have intensity 60 cos θ *cd* in a direction θ from the normal. The brightness of a blackbody is the same in all directions, and its value for the 2,046 K surface would be 60×10^4 cd/m^2.

25·2. Some Theorems in Photometry

In simple situations several inter-relations hold good between the quantities defined above. We will see some of them which are used widely.

(*a*) *Relation* $E=I/(r^2 \cos\theta)$

If the dimensions of the source are much smaller than the distance of the observer, then the source is called a 'point source'. In such cases, the solid angle $\Delta\Omega$ subtended by a given surface o area $\triangle S$ at the source is defineable : We have $\Delta\Omega = \triangle S \cos\theta / r^2$, where θ is the angle between the ray and the normal to the surface (Fig. 25·1). If the source has intensity I *cd* in the concerned direction, the flux received is $Id\Omega$ *lm*, and illumination is

$$E = \frac{Id\Omega}{dS} = \frac{IdS\cos\theta}{r^2 dS} = \frac{I\cos\theta^*}{r^2} \qquad \ldots(25{\cdot}1)$$

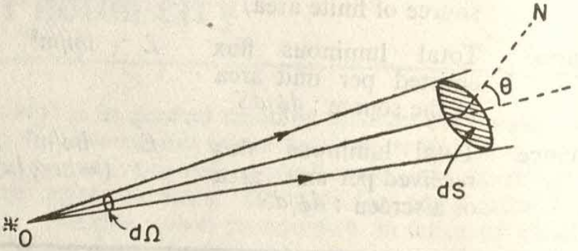


Fig. 25·1.
Illumination by a point source.

This is one of the *Lambert relations.* It combines the inverse square law with $\cos\theta$ dependence. *But it is useful only if the source can be treated as a point source.*

(*b*) *Relation* $L = \pi B$

For a surface of uniform brightness B, the luminance L is given by πB. The elementary solid angle for the hollow cone between θ and $\theta + d\theta$ from the normal to a surface is $2\pi r \sin\theta .\ r d\theta \div r^2 = 2\pi \sin\theta\, d\theta$. The brightness is defined by

$$B = \frac{\Delta I}{\triangle S \cos\theta}$$

Hence the luminous flux in the hollow cone due to area $\triangle S$ is $B \triangle S \cos\theta .\ 2\pi \sin\theta\, d\theta$. Integrating this over the whole available solid angle gives $L . \triangle S$. Thus

$$L \triangle S = \int_0^{\pi/2} B \triangle S \cos\theta\ 2\pi \sin\theta\, d\theta$$

If B is taken as independent of θ, we get the result

$$L = \pi B \qquad \ldots(25{\cdot}3)$$

This is another *Lambert relation.* One consequence of this is that luminosity of the blackbody surface used for defining the Candela is $6\pi \times 10^5\ lm/m^2$. [check it].

* A dimensional examination is of interest.

$$[E] = [I]\,[r^{-2}] = [lm\ L_x^2 L_y^{-2}]\ [L_x^{-2}] = [lm\ L_y^{-2}]$$

This tallies with Table 25·1. With equal validity we could put

$$[E] = [cd]\ [L_x^{-2}]$$

But this is not the same as the dimension of B, which is $[cd]\ [L_y^{-2}]$

(c) *Brightness of a surface is independent of the distance of the viewer*

This is because the illuminance falls as $1/r^2$, and the solid angle subtended at the eye by a given area of the surface also falls as $1/r^2$. Hence the flux received *per unit solid angle* remains unchanged. Of course, this assumes that there is no attenuation in the medium.

(d) *Relation* $E=\pi L$

The illuminance at the centre of a flat surface under a hemispherical cover of uniform luminance L is given by πL. An elementary ring on the hemisphere defined by θ and $\theta+d\theta$ has area $2\pi r \sin\theta.\, r d\theta$ and hence intensity $L.2\pi r^2 \sin\theta\, d\theta$. Its contribution to the illuminance at the centre is $\cos\theta/r^2$ times this (see Eq. 25.1). That gives

$$dE=L.2\pi \sin\theta \cos\theta\, d\theta \qquad \text{...(25.4)*}$$

Integration from $\theta=0$ to $\theta=\pi/2$ gives

$$E=\pi L$$

The result is independent of r.

The illuminance over the surface due to a parallel infinite surface of luminance L is also πL. The proof of this we will leave for the reader. The result is independent of the distance between the two surfaces.

(e) *Relation* $I=B.A$

For a surface of uniform brightness B, the intensity for a distant observer is $B.A$, where A is the area of the source projected on a plane transverse to the line of slight. This follows directly from Eq. (25·2)

$$dI=B\, dS \cos\theta=BdS'$$

$$I=B\int dS'=B.A. \qquad \text{...(25·5)}$$

The relation has special interest in the viewing of stars (see example 25·3).

(f) *No image can have brightness greater than that of the source*

Fig. 25·2 shows the geometry in image formation, with areas S, S' and solid angles Ω, Ω' shown, besides u and v.

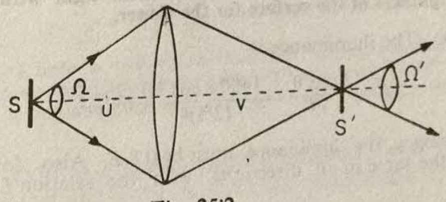


Fig. 25·2.

To discuss brightness of images

---

* It is of interest that π in this equation is dimensionless, where as π in Eq. (25·3) has the dimensions $(L_y^2 L_x^{-2})$. The reader may follow the steps of derivation to see this difference.

Then we have from geometry

$$S'=\frac{v^2}{u^2}S\,; \qquad \Omega'=\frac{u^2}{v^2}\Omega$$

If we consider a source of uniform brightness B, then the luminous flux collected by the lens is $B.S.\Omega$ and this spreads over area S', so that *illuminance* E on the image plane is

$$E=B\Omega\frac{S}{S'}=B\frac{A}{u^2}\cdot\frac{}{v}=B\frac{A}{v^2} \qquad \ldots(25\cdot6)$$

where A is the area of the lens. Thus E can be increased by having larger lens area and smaller focal length (to give lower v)

But consider *brightness* B' of the image. The flux collected by the lens emerges in a solid angle Ω'. Hence the flux per unit area of the image per unit solid angle (which is B') comes to

$$B'=\frac{BS\Omega}{S'\Omega'}=B\left(\frac{S}{S'}\right)\left(\frac{\Omega}{\Omega'}\right)=B\frac{u^2}{v^2}\cdot\frac{v^2}{u^2}=B \qquad \ldots(25\cdot7)$$

Thus the best we can get is $B'=B$; in practice losses due to various effects give $B'<B$.

Example 25·1. A source of light has intensity depending on θ according to $I=I_0\cos(\theta/2)$ Deduce the total flux emitted by it, and the averaged intensity as a point source.

Solution. The solid angle between θ and $\theta+d\theta$ is $2\pi\sin\theta\,d\theta$.

Hence $\quad d\phi=I\,2\pi\sin\theta\,d\theta=I_0\,2\pi\sin\theta\cos\tfrac{1}{2}\theta\,d\theta$

Integrating from $\theta=0$ to $\theta=\pi$ (point source)

$$\phi=2\pi I_0\left[-\frac{4}{3}\cos^3\frac{\theta}{2}\right]_0^{\pi}=\frac{8\pi}{3}I_0$$

$$I_{av}=\frac{\phi}{4\pi}=\frac{2}{3}I_0$$

⊙

Example 25·2. A source P and an observer Q are placed at angles 30° and 60° respectively from the normal to a surface. The source has intensity 1400 cd in the concerned direction, and its distance from the surface is 2·5 m. The surface gives diffuse scattering of the incident light with 80% efficiency, Deduce the brightness of the surface for the viewer.

Solution. The illuminance is

$$E=\frac{I\cos\theta}{r^2}=\frac{1400\times\cos 30}{(2\cdot5)^2}\ lm/m^2$$

With 80% efficiency, the luminance must be $0\cdot8\,E$. Also, for an ideal diffuse scatterer, B is the same in all directions, with the relation $L=\pi B$ (Eq. 25·3). That gives

$$B=\frac{L}{\pi}=\frac{0\cdot8E}{\pi}=\frac{0\cdot8\times1400\times\cos 30}{\pi(2\cdot5)^2}\ cd/m^2.$$

(Notice that angle 60° is irrelevant for an ideal diffuse scatterer).

⊙

Example 25·3. The angular diameter of a star is negligibly small compared to the diffraction spread λ/a due to viewing with an instrument of aperture diameter a. If the effective brightness of the star is 10^{-4} times that of the background sky in viewing with the naked eye, what will the relation be in viewing through a telescope of 200 times the diameter of the pupil of the eye ?

Solution. We will use subscripts s and o for the star and the background sky, and primes ′ and ″ for viewing with the eye and the telescope respectively. We have the general relations.

$$I=BA=B'A'=B''A''$$

and $$A'=r^2\Omega$$

where B is the brightness, A is the transverse area, r is the distance, and Ω the solid angle. For the star, Ω arises due primarily to diffraction, so that

$$\Omega_s'=\pi\,(\lambda/a')^2\,; \qquad \Omega_s''=\pi\,(\lambda/a'')^2$$

Substitutions lead to

$$B_s''=B_s'\,(a''/a')^2$$

In the case of the sky, the intrinsic $\Omega_\circ$ far exceeds the contributions due to diffraction, and hence $\Omega_o''=\Omega_o'=\Omega_o$. That means

$$B_\circ''=B_\circ'$$

Hence $$(B_s''/B_\circ'')=(B_s/B_\circ')\,(a''/a')^2$$
$$=(10^{-4})\,(200)^2=4$$

Thus, in the telescope the star brightness is 4 times the sky brightness

(Note : It is incorrect to take the effect as due to larger light gathering by a telescope lens. That effect is common for the star and the background light. The effect is because of smaller diffraction spread in star image as a increases)

Example 25·4. A typical coiled-coil filament lamp of 100 cd intensity has a filament of apparent length 2·5 cm and diameter 0·30 cm. Deduce the total flux emission, luminosity and brightness, assuming isotropic character of emission.

Solution

$$\text{Total flux}=\ Id\Omega=I.4\pi=400\pi\ lm$$
$$\text{Area of cross-section}=2{\cdot}5\times0{\cdot}30\times10^{-4}m^2=7{\cdot}5\times10^{-5}\ m^2$$

Luminosity ($\mathrm{L}_{average}$)

$$=\frac{\phi}{S}=\frac{400\pi\ lm}{7{\cdot}5\times10^{-5}\ m^2}=5{\cdot}3\pi\times10^6\ lm/m^2$$

Brightness ($\mathrm{B}_{average}$)

$$=\frac{I}{S}=\frac{400\ cd}{7{\cdot}5\times10^{-5}}=5{\cdot}3\pi\times10^6\ cd/m^2$$

(Note : L here is $\sim$9 times that for the surface defining the candela. This is not because of higher temperature—T is of same order—but is because of coiled coil. The actual coil is much longer than that given as apparent length here).

25·3. Some Typical Intensity and Illuminance Values

For comfortable work in different situations the illuminations recommended are as follows :

Home, libraries, office, etc. : 100-300 lx
sewing and finer work : 500-1000 lx
operation theatre table : 5000-8000 lx

The accomodation provided by nature in the eye is able to take care if there is excess illumination ; but in poor illumination one has less than the optimum efficiency of work.

A typical 100 W bulb of incandescent filament type has intensity ~200 *cd* in the forward direction, so that the illuminance 2*m* below it would be ~50 lx due to direct light. Multiple scattering from walls of the room raise this to 2 to 5-fold, depending on the wall colour and room size.

Sunlight at mid-day produces an illuminance about 10^5 lx. In full-moon at mid-night it is about 0·25 lx. That we can at least see, if not read, under such a large range of illuminations is remarkable.

A sodium vapour lamp has much larger *intensity* than an incandescent filament lamp ; but the filament could be much brighter than the sodium vapour lamp, because the area is very much smaller the second case.

Any illuminated surface acts as a secondary source of luminous flux, defined by brightness *B* and luminance *L*, just as for a primary source. Typically, a matt surface coated with MgO will have luminance *L* about 95% of the illuminance *E*, and also have *B* equal in all directions. A mirror reflector, on the other hand, will have *B* zero in all directions except the range of specular reflection and *L* around 80-98% of *E*, depending on the quality of the reflecting polish.

Example 25·5. A lamp of I=400 cd hangs 1·5 m above the centre of a table of dimensions 2·0 m×2·0 m. Deduce the illumination at a diagonal end of the table. If a paper placed there gives diffuse scattering with 60% efficiency, deduce the iuminance of the paper and also the brightness.

Solution. For the diagonal end,

$$r^2=[(1\cdot5)^2+(1^2+1^2)]=4\cdot25\ \text{m}^2;$$

$$\cos\theta=\frac{1\cdot5}{r}$$

$$\text{Illuminance } E=\frac{I\cos\theta}{r^2}=\frac{400\times1\cdot5}{(4\cdot25)^{3/2}}\approx70\ \text{lx}$$

$$\text{Luminance } L=0\cdot60\times E\approx42\ \text{lm/m}^2$$

$$\text{Brightness } B=L/\pi\approx14\ \text{cd/m}^2$$

⊙

25·4. Luminosity and Luminous Efficiency of a Source

A primary source of light consumes energy at the rate of *X* watt. Out of this a certain part, *X′* watt, goes out as radiant enery, the rest is consumed otherwise. This *X′* watt creates a total luminous flux ϕ *lm*. Then the ratio (ϕ/X') *lm*/*W* is called the luminosity of the source, and the ratio (ϕ/X) *lm*/*W* is called the *luminous efficiency* of the source. Not unoften, the distinction between *X* and *X′* is omitted and one writes

$$\eta = \frac{\text{Luminous flux } (lm)}{\text{Power concerned } (W)} \qquad \text{...(25·8)}$$

where η is used for both luminosity and luminous efficiency.

The value of η defined as ϕ/X', depends on the spectral distribution of the radiation. Obviously, if all the radiation falls outside the visible range ($\sim$4000 to $\sim$7000 A), then $\phi=0$ and hence $\eta=0$. Within the visible range, ϕ/X' is wavelength-dependent, being the maximum for the green region $\lambda \sim 5550\ A°$, and falling off with λ both ways. This maximum value is

$$\eta_{max} = 683\ lm/W \text{ at } \lambda = 5550\ A$$

The number 683 comes from the way 1 candela is defined. Fig. 25·3 shows the relative spectral sensitivity, K_λ, of the human eye.

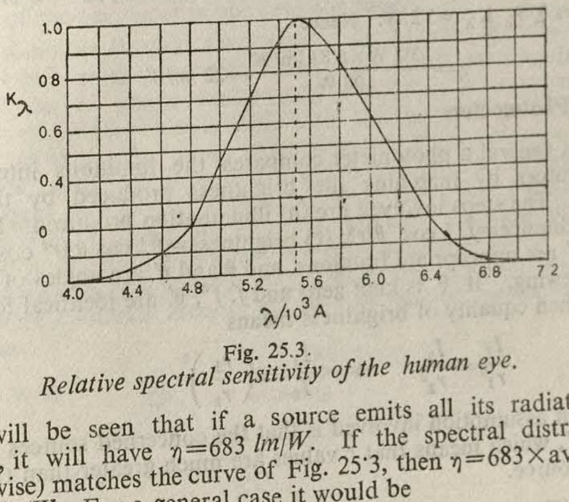


Fig. 25.3.

Relative spectral sensitivity of the human eye.

It will be seen that if a source emits all its radiation at $\lambda=5550A$, it will have $\eta=683\ lm/W$. If the spectral distribution (intensitywise) matches the curve of Fig. 25·3, then $\eta=683\times$averaged $K_\lambda \approx 200\ lm/W$. For a general case it would be

$$\eta = \frac{\int K_\lambda P_\lambda d\lambda}{\int P_\lambda d\lambda} \times 683\ lm/W \qquad \text{...(25·9)}$$

where P_λ is the radiated power at λ. This is for η defined by ϕ/X'. The realistic η is defined by ϕ/X and it will be lower, depending upon the power $(X\text{-}X')$ consumed otherwise by the concerned system.

Sunlight happens to have its radiant energy maximum close to $5550A°$, and it appears that nature has adapted the human eye to match its sensitivity curve to secure the best from sunlight. For other sources the typical η values are as follows :

incandescent lamp	:	14 lm/W
fluorescent lamp	:	43 lm/W
Halogen lamp	:	58 lm/W

The values mean that a 100 *W* incandescent lamp would emit ~1400 *lm* flux. Considered isotropic, its intensity would be $1400/4\pi \approx 120$ cd. With a 'shade', the *I* value in the front direction may reach ~400 cd ; from a 2*m* distance it would create illuminance of ~100 *lx*.

Example 25·6. A 100 W lamp may be treated as emitting radiations at different wavelengths as listed below ;

λ/1000 A=4·8	**5·2**	**5·6**	**6·0**	**6·4**	**6·8**	**7·2**
P_λ/Watt=0·4	**0·8**	**1·2**	**1·6**	**2·0**	**2·8**	**3·0**

The rest of the power is spent otherwise. Using the sensitivity curve of the human eye (Fig. 25·3), and given η_{max}=683 lm/W, deduce the η of the lamp.

Solution. We list the corresponding K_λ values and hence $P_\lambda\ K_\lambda$

K_λ =	0·20	0·71	0·95	0·60	0·16	0·04	0 00
$P_\lambda\ K_\lambda$ =	0·08	0·57	1·14	0 96	0·32	0·11	0·00

That gives $\Sigma\ P_\lambda\ K_\lambda$ =3·2 W. Hence

$$\eta = \frac{3{\cdot}2\ \mathrm{W} \times 6{\cdot}83\ \mathrm{lm/W}}{100\ \mathrm{W}} = 22\ lm/W$$

⊙

25·5. Photometers

In general a photometer compares the luminous intensities of two sources by matching the brightness produced by them on a surface. The steps involved are (*a*) illumination produced$=I \cos\theta/r^2$; (*b*) luminance$=f.\ I\cos\theta/r^2$, (*c*) brightness$=ff'\cos\theta/(r^2\cos\theta')$. Here f and f' are appropriate fractions, and θ and θ' are angles of incidence and viewing. If θ is kept zero and f, f', θ' are identical for the two cases, then equality of brightness means

$$\frac{I_1}{r_1^2}=\frac{I_2}{r_2^2} \Rightarrow \frac{I_1}{I_2}=\left(\frac{r_1}{r_2}\right)^2 \qquad (25{\cdot}9)$$

An assumption involved is that the concerned sources are 'point sources', which means that *r* values are much greater than the size of either source.

Different photometers use different methods of matching the brightness. We describe some of them.

(*a*) *Bunsen's Grease-Spot Photometer*. It uses a mat-white paper (like a filter paper) with a central greased spot. The ungreased part gives diffuse scattering of light falling on the same face, while the greased part gives diffuse scattering of light falling on the opposite face. Fig. 25.4 shows the arrangement. If the distances of the sources are adjusted to make the grease spot vanish in the background of the ungreased part, Eq. (25.9) gives I_1/I_2. If we assume that the ungreased part scatters a fraction *a*, and the greased part scatters fraction *b* and *c* of the light falling on the same and the opposite face, respectively, then the brightness of face 1 is

$(I_1/r_1^2)\ a$ for the ungreased part

$(I_1/r_1^2)\ b+(I_2/r_2^2)\ c$ for the greased part ...(25·10)

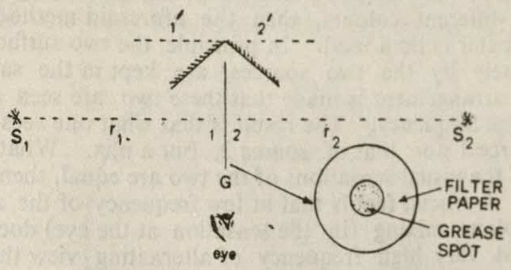


Fig. 25·4.
The grease-spot photometer

Equality of brightness will still lead to Eq. 25.9 if $b+c=a$ (*i.e.*, there is no loss by absorption). In practice, the grease spot never fully vanishes in the background, and what one has to adjust is the 'identical view' of the two faces of the detector paper.

(*b*) *Lummer-Brodhun Photometer*. In this also the two sides of a matt white surface (a thick one, this time) are illuminated by light from the two sources under study (Fig. 25·5), but comparison of brightness of the two faces is made with the help of a 'photometer cube' *C*. In this 'cube' an air space separates the hypotenuse faces of two prisms, but a central portion is sealed with canada balsam, which is transparent, like glass. Thus an observer views face 1 through the central portion of *C* and face 2 through the outer portion of *C*. In effect, this is like setting $b=0$ and $c=a$ in relations 25·10. Because total internal reflection gives 100% reflection, these conditions hold well, assuming that the faces of all the right-angled prisms are kept very clean.

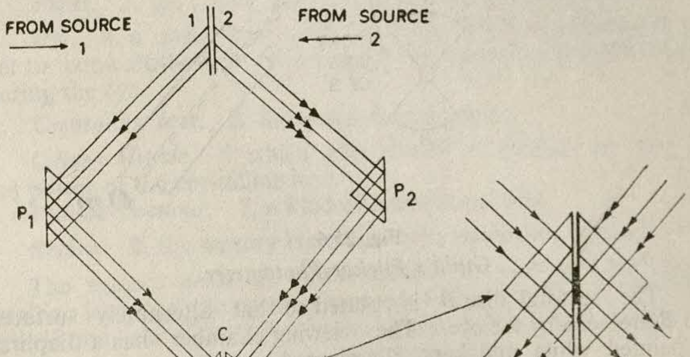


Fig. 25·5.
Lummer-Brodhun photometer, using total reflection prisms P_1 and P_2 and the Lummer-Brodhun 'cube' C. Inset shows how the rays from sources 1 & 2 pass through the central and outer parts of C.

(*c*) *Flicker Photometers.* If the sources under comparison give out light of different colours, then the aforesaid methods fail. A flicker photometer is then used. In principle, the two surfaces, illuminated separately by the two sources, are kept in the same line of sight, but an arrangement is made that these two are seen alternately at a fairly large frequency. The result is that what one sees is neither colour of source 1 nor that of source 2, but a mix. What is important is that if the visual sensations of the two are equal, then no flicker is observed. A crucial fact is that at low frequency of the alternating view, the colour mixing (in the sensation at the eye) does not take place, and at very high frequency of alternating view the flicker is not observable even when brightnesses are unequal (due to persistence of vision). So a careful balance is needed. One starts from low rate and goes on to higher ones to see if the colour distinction becomes absent; and then one adjusts the distances till the flicker is eliminated.

Fig. 25.6 shows the schematic of *Guild's flicker photometer.* Surface A is illuminated with source 1 and seen at 45° from its normal. Surface B is a sector disc illuminated with source 2 and seen again at 45° from the normal. As the sector disc is rotated, the eye alternately views A and B. If both the surfaces have identical diffuse coating (say, of MgO), which is neutral for colours, then absence of flicker means $I_1/r_1^2=I_2/r_2^2$.

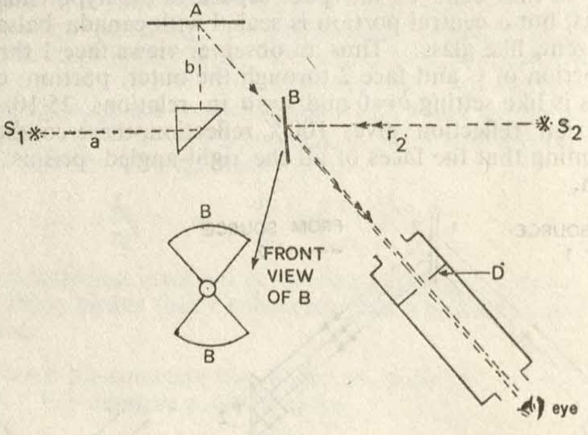


Fig. 25·6.
Guild's Flicker Photometer

The sectoral disc B is rotated so that alternately surfaces A and B are seen by the eye. The viewing chamber has a diaphragm D painted white and kept illuminated to serve as a comfortable background against which A/B is seen ($r_1=a+b$).

(*d*) *Ulbright's Integral Photometer.* The luminous intensity of a source is often direction-dependent. One may therefore measure I in different orientations and then find the averaged I over the total solid angle. But Ulbright's photometer does the integration in another way. Fig. 25.7 shows the source S placed inside a hollow sphere, whose inside walls are perfectly diffusing, mat painted typically with MgO. Due to multiple reflections within, the radiation inside becomes isotropic. If a window *W* is placed on the wall, the output flux there is equal to the total flux emitted by the source. Comparing the brightness of an opal glass placed at this window with that produced by a standard source from a measured distance gives the desired result.

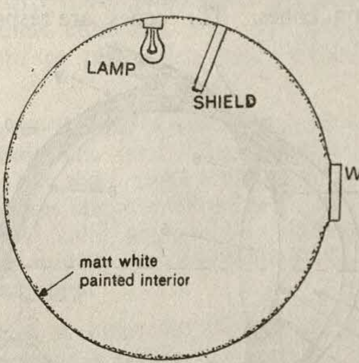


Fig. 25·7.

Ulbright's integral photometer G. is gate for introducing the source S; P is a sheild to prevent direct light from reacting W.

25.6. The Human Eye

A sketch of the human eye is shown in Fig. 25.8. The different optical parts are as follows :

Cornea. 1, a thin transparent front opening.

Aqueous Humour. 2, a watery fluid between the cornea and the crystalline lens.

Pupil. 3, the central opening of the crystalline lens front.

Iris. 4, a pigmented muscular ring which can expand or contract to control the size of the pupil and hence the amount of light entering the eye.

Crystalline lens. 5, the image forming lens.

Ciliary Muscle. 6, which can stretch or slacken to vary the focal power of the crystalline lens.

Vitreous humour. 7, a kind of transparent jelly.

Retina. 8, the sensory layer on which images are formed.

The system made up of the cornea, the aqueous humour, the crystalline lens, and the vitreous humour add up to an effective focal length 17 mm; the ciliary muscle can change it by a small margin.

The major mechanical parts of the eye are :

Scalera. 9, which encloses the eyeball, protects the inside, and gives it the needed stiffness.

Vascular layer. 10, which support the rods and cones behind the retina.

Finally, the parts which convert the physical image into the vsual sensations of intensity and colour. The *retina*, the shape of a hemisphere, contains light receptors, called *rods* and *cones*. These are light-sensitive cells. In the human eye there are ~ 125 million rods and 6.5 million cones. The cones are responsible for colour

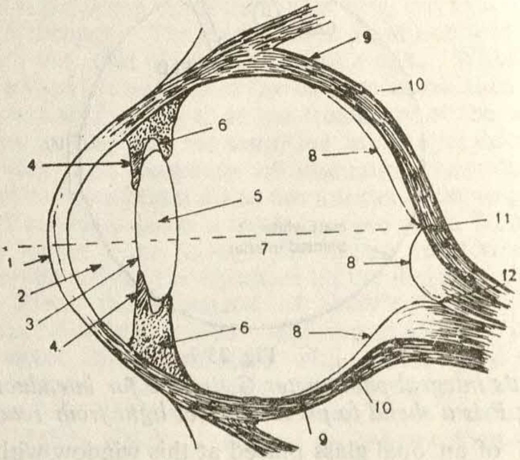


Fig. 25·8.

The Human Eye.

vision and are concentrated at the depression in the retina, called *fovca centralis*, 11, which falls on the optic axis. The remaining areas are occupied mainly by the rods, with the exception of the junction of cells of the retina with the optic nerve, 12; there are no rods or cones here, so that this junction forms the blind spot.

In rods the light causes the rhodosin [a compound of a form of vitamin A (retinen) and the protein of the retina (opsin)] to change from the cis— to trans— isomer. Thereby each rod generates a *nerve impulse*, which is conveyed by the *optic nerve* to the brain. In the process, the energy is drawn from that stored already by the receptors; the light only triggers the reaction. That is the reason why each rod is capable of responding to a photon.

Rods do not perceive colour. That is the function of the cones. Hence colours are preceived only if the image falls into the central area of the retina. Probably there are at least three kinds of cones, perceiving green, red and blue. Intermediate colours are seen depending on the relative stimulation. The brain perceives the colours on the basis of different trains of pulses from the cones. However, there are many unexplained features and a consistent theory of colour vision is still lacking.

PROBLEMS

25.1. A 60W lamp has intensity 100 cd and is isotropic. Deduce (*a*) the luminous efficiency, (b) illuminance on a surface placed 4.0 m away with inclination $\theta=30°$. If the surface is a diffuse scatterer with 60% efficiency, deduce its luminance, and brightness.

25.2 A 100 W lamp has uniform intensity 250 cd in a cone of semiangle 60° and zero intensity outside. Deduce (*a*) the luminous flux, (*b*) the luminous efficiency, and (*c*) brightness of a perfect diffuse scatterer surface placed 5 m away normal to the incident beam.

25.3 In a street lighting system, the poles are 12 m apart and the lamps 5.0 m above the road. The lamps are designed to give uniform intensity over a solid angle 1.5π sr and zero outside. If the averaged intensity of the lamps is 600 cd each, deduce (*a*) the intensity within the 1.5 π *sr* solid angle cone, (*b*) the illuminance at the bottom of each lamp, and (*c*) the illuminance at the mid-point between the bases of adjacent poles.

25.4. (*a*) *Phot*, is a unit for illuminance equal to 1 lm/cm^2. Deduce the relation between a phot and a lux.

(*b*) *lambert* is a unit of brightness equal to 1 cd/cm^2. Deduce the relation between a lambert and 1 cd/m^2.

(*c*) *metre-candle* is a unit for illuminance. Deduce in metre-candles the illuminance for a surface placed 2 m from a source of intensity 100 cd and normal to the beam.

25.5. (*a*) A chamber encloses *diffuse* radiations. If the flux across unit area anywhere, considered in all directions, is ϕ, show that the flux component along one particular direction is $(1/4)\phi$.

(*b*) If the above chamber has a long cylindrical window of cross-sectional area S′ and a source of average spherical intensity *I cd*, deduce the value of ϕ. How will the result change if a window of same area S′ is cut right on the chamber face ?

25·6. A collimated beam is created by placing a source at the focus of a convergent lens of $f=25$ cm and aperture diameter 4·0 cm. If the source is taken as a sphere of diameter 3.0 mm, deduce the cross-sectional diameter of the beam at distances (*i*) 10 cm, (*ii*) 1.0 m, (*iii*) 10 m from the lens, ignoring diffraction effect. If the source has intensity 60 cd in the concerned range, deduce the illuminance in the beam at the three given distances.

25·7. Comment critically on the following :

(*a*) The contrast of star brightness relative to the background sky brightness improves when telescopes of large apertures are used ; but this is *not* because of larger collection of luminous flux.

(*b*) In a given image-forming system, the narrowing of lens aperture does *not* affect brightness of the image.

(*c*) A sodium vapour lamp has much larger *intensity* than a typical incandescent filament lamp, but in *brightness* the latter could greater.

25·8. Two sources of light of intensities I and $6I$ are placed 200 cm apart. At what place should a screen be placed on the line joining the sources, so that illumination due to the two sources may be equal on the screen.

25·9. A photometric balance is obtained between two sources P and Q when P is 58·0 cm from the photometer. When a plane glass plate is introduced in the beam from P, with its face normal to the beam, the balance is restored when P is shifted by 2·6 cm. Deduce the percentage loss of light by reflection at each face of the plate. What do you expect if the plate were inclined from the beam at an increasing angle ?

25·10. The angular diameter of the moon, as seen from the earth is 0·008 radian, and the moon's surface re-emits 20% of the light incident on it from the sun. Deduce the ratio of illumination on the earth in full moonlight to that in full sunlight.

25·11. Deduce the luminance of the sun, given that its disc subtends an angle 30′ arc and the illuminance due to it at earth's surface at mid-day is 10^5 lux. An image of the sun is formed by a lens of diameter 8 cm and focal length 30 cm; deduce the illuminance of the image.

25·12. For the human eye, consider the pupil diameter $\sim$3 mm, mean wavelength $\sim$6000A, and focal length $\sim$20 mm. Compute the area of the image of an ideal point source at infinity as determined by diffraction. If the areal density of the cones in the retina is designed by nature to use the optimum available resolution, estimate the areal density of the cones on the retina.

Answers to Numerical Problems

2·2 Simi-angle 30′

2·3 (*i*) 35·9 cm, 33·1 cm ; (*ii*) −1·9 ; (*iii*) 3·6.

2·4 H at 51·0 cm, H' at 18·8 cm. $f=20{\cdot}4$ cm, $f'=-30{\cdot}7$ cm.

2·5 Image sizes −7·19, −7·24, −7·63 mm for the three zones.

3·5 $f'=7{\cdot}5$, 9·0, 11·2, 15·0, 22·5 cm.

3·9 $f'=-f=2{\cdot}67$ cm ; H and H' at centre of the sphere.

3·10 Focal points F' at 4·0 cm from curved face and F at 3·5 cm from flat face, both outside. H' on the curved face, H at 1·7 cm from the flat face, inside.

3.11 $\alpha'=-2{\cdot}14$ cm, $\alpha=+1{\cdot}42$ cm (both H, H' inside the lens), $f'=12{\cdot}86$ cm, $f=-12{\cdot}86$ cm.

3·12 $\alpha'=+2{\cdot}35$ cm, $\alpha=+3{\cdot}53$ cm (both H, H' outside the lens), $f'=52{\cdot}9$ cm, $f=-52{\cdot}9$ cm.

3·13 $\alpha'=+40$ cm, $\alpha=+80$ cm (both H, H' outside the pair), $f'=-100$ cm, $f=+100$ cm.

3·14 $m=-0{\cdot}58$.

3·15 $f'_1=f'_2=t=6$ cm.

3·17 $f'=10{\cdot}9$ cm.

4·2 (*i*) 1·64 cm, (*ii*) 3·28 cm.

4·5 Crown $f=13{\cdot}12$ cm, flint $f'=-19{\cdot}52$ cm. $G=0{\cdot}1469\ \text{cm}^{-1}$ and $-0{\cdot}0885\ \text{cm}^{-1}$ respectively.

4·6 (*i*) G (Crown)$=0{\cdot}4046\ \text{cm}^{-1}$, $G(EDF)=-{\cdot}01570\ \text{cm}^{-1}$. (*ii*) $G=$(Crown)$={\cdot}03452\ \text{cm}^{-1}$, $G(EDF)=-{\cdot}01133\ \text{cm}^{-1}$.

4·7 For crown lens $R_1=-R_2=+27{\cdot}06$ cm, for flint lens $R_1=-27{\cdot}06$ cm, $R_2=-135{\cdot}1$ cm.

4·10 (*i*) 1°18′, (*ii*) 6°8′.

4.11 $\Delta m=0{\cdot}32$.

5·1 113 cm.

5.2 0·061.

5·3 25 : 44 : 144. Same ratio for light.

5·5 $f_1=10$ cm, $f_2=10/3$ cm, $t=20/3$ cm.

5·7 ·485 cm, 13·20 cm below the flat face. Areal magnification 7·41.

5·8 12 cm.

5·9 8·75×

5·11 Image plane concave towards the object, with radius of curvature$\sim$10 cm.

6·1 0·3 rab ; 0·5 rad.

6·7 $f_1'=f_2'=5{\cdot}6$ cm, $t=3{\cdot}73$ cm.

6·9 $f_1'=4{\cdot}9$ cm, $f_2'=5{\cdot}6$ cm, $t=4{\cdot}4$ cm, $L_2E=0{\cdot}8$ cm, $r=5{\cdot}7$.

6·12. $2{\cdot}5\times10^{-4}$ cm.

7·2 (*i*) 6·0 Km, (*ii*) 4500 rev/min.

7·3 4 m spread.

7·7 $18{\cdot}25\times10^{-8}$ m s^{-1}.

8·3 $c>2{\cdot}2\times10^{7}$ m s^{-1}.

8·7 $1 : 10^{-2} : 6\times10^{-3}$ and $1 : 10^{-2} : 60$.

8·8 $2{\cdot}5\times10^{-9}$ s^{-1}.

8·9 $7{\cdot}7\times10^{-2}$ V m^{-1}.

8·11 $E^\circ=56$ Vm^{-1}; $D^\circ=5{\cdot}0\times10^{-10}$ coul. m^{-2}; $H^\circ=0{\cdot}15$ amp. m^{-1}; $B^\circ=1{\cdot}9\times10^{-7}$ V m^{-2} s.

8·12 $E^\circ=2{\cdot}5\times10^{-5}$ Vm^{-1}; $H^\circ=6{\cdot}6\times10^{-8}$ amp m^{-1}.

9·2 $A_C=1{\cdot}71$, $A_S=-4{\cdot}70$, $\omega=1000$ units; $a=0{\cdot}85+2{\cdot}35\,i$.
(*b*) $A=10$ units, $\phi=\tan^{-1}(3/4)$; $A_C=8$, $A_S=-6$ units.

9·3 117 units, +0.92 rad, 1/5 units of time, 80.

9·4 $a=2{\cdot}0\times10^{-2}$V/m, $\omega=3{\cdot}0\times10^{14}$ s^{-1}, $k=0{\cdot}005$ s^{-1}.
$|E^2|_o=4{\cdot}0\times10^{-4}$ V^2 m^{-2}; $I=0{\cdot}53\times10^{-6}$ Jm^{-2} s^{-1}.

9·8 $0{\cdot}914a$, $0{\cdot}073a$, $0{\cdot}011a$.

9·9 $3{\cdot}0\times10^{-3}$ Vm^{-1}, $1{\cdot}34\times10^{13}$ *Hz*, $2{\cdot}24\times10^{-5}$ m, $3{\cdot}0\times10^{8}$ m s^{-1}.

9·12 (*i*) $1{\cdot}958\times10^{8}$ m s^{-1}, (*ii*) $1{\cdot}868\times10^{8}$ m s^{-1},
(*iii*) $-1{\cdot}534\times10^{6}$ m s^{-1}, (*iv*) 5·9.

10·5 0·51; 0·39.

10·6 $1{\cdot}875\times10^{8}$ m/s ; 3413 A°.

10.7 56% loss.

11.2 $A/a=2$, 1·932, 1·732, 1·414, 1·000, 0·518, 0·518, 1·000......

11·3 $A/a=5{\cdot}738$, 3·863, 0, 1·414.

11·4 0·820.

11·9 $0{\cdot}894\pi$.

11·10 20 cm.

11·12 66·9, 83·8.

11·13 0·246 cm, 0·164 mm.

11·14 0·82 mm towards *A*.

11·15 1·000379.

11·16 491·5 (a minimum), 530·7, 596·1.

11·17 $\sin^{-1} 0{\cdot}04n$ ($n=1, 2, \ldots\ldots$); 20 sin θ Km altitude.

11·18 10 and 4 nearly.

12.3 $(1{\cdot}77 \times n)$ micron, $n=1, 2 \ldots\ldots$

12·4 $5{\cdot}89 \times 10^{-6}$ cm.

12·7 0·94 mm.

12·8 ·0589 mm.

12·9 $1{\cdot}02 \times 10^{-5}$ per °C.

12·10 16·6, 49·8, 83·0......micron.

12·12 0·764 cm, 0·935 cm, $5{\cdot}05 \times 10^{-5}$ cm.

12·13 0·139 cm^2.

12·15 $5{\cdot}78 \times 10^{-5}$ cm.

12·16 Fringes emerge out, visibility becoming minimum 5 times and maximum at 4 intervening positions.

12·17 (i) $2{\cdot}33 \times 10^{-3}$ rad, (ii) 20·3, (iii) 127.

12·18 1·000289.

12·19 ·035, ·018, ·0080, ·0016 part of fringe separation.

13·2 First case : $a\sqrt{2}$, 0, $a\sqrt{2}$, 0. Second case : 1·26a, 0·33a, 1·12a, 0·40a.

13·3 $4{\cdot}7 \times 10^{-3}$ cm^2. (i) $5{\cdot}5 \times 10^{-2}$ cm. (ii) $6{\cdot}7 \times 10^{-2}$ cm.

13·7 $(1000/n)$ metre ; $(10/n)$ metre ; $n=1, 2, 3 \ldots\ldots$

13·8 (i) 35·7 cm ; (ii) 12·1, 6·6 cm ; (iii) 1/9, 1/25.

13·9 77 cm ; 1·86 mm.

13·10 250 cm, 0·76 mm, 2·36 mm.

13·12 Maxima : 2·1, 3·63, 4·70 mm. Minima : 2·97, 4·20 mm.

14·1 $>0{\cdot}9$.

14·3 $\pm\sin^{-1}\frac{3}{15}$ and $\pm\sin^{-1}\frac{1}{2}$ from the normal atong length and breadth respectively.

14·5 Initially, minima at ±14°30′, 30°, 48°30′ and 90°. At $a=5{\cdot}0$ cm minima at ±30° and 90°. At $a=2{\cdot}5$ cm, minima at ±90°. As a further decreases, the intensity spreads more from 0° towards ±90°.

14·6 due to slit : 10^{-4} rad ; due to diffraction : $\sim 0{\cdot}5 \times 10^{-4}$ rad.

14·7 0·5 cm.

14·10 distribution defined by cones of semi-angle θ with line of oscillators as the axis. Minima at $\theta=\sin^{-1}(n/10)$, with $n=0, 1, 2, 3 \ldots 10$

14·13 (i) 3 orders, (ii) 3 orders, (iii) 7 orders. Second order of 7500A° and fourth order of 3750A°.

14·14 41·7μ.

14·17 ~13°.

14·18 5147·1A°, 5158·1 A°.

14·19 0·432 mm/A°.

15·1 (i) 250,000 and 5,000 ; (ii) 0·80 cm^{-1} and 0·02 cm^{-1}

15·2 $\sim$270 m ; $\sim$50 m.

15·3 $2{\cdot}4\times10^{-4}$ rad, $1{\cdot}9\times10^{-4}$ rad.

15·4 1/1000 rad. N-S ; 1/250 rad. E-W.

15·5 $1{\cdot}25\times10^{-4}$ rad/A° ; 50,000 ; 0·1 A° ; 0·4 cm^{-1}

15·6 Obsd. line becomes twice in width ; thrice in width.

15·7 (i) 53°, (ii) $3{\cdot}08\times10^{-4}$ rad/A°, (iii) $7{\cdot}52\times10^{-4}$ rad, (iv) 24×10^{-8} A.

15.8 0·16 mm/A° ; 0·083 A° ; 0·33 cm^{-1}.

15·9 0·41 A°

15·10 1·6A°.

15·11 diameters of stars are less than $\sim10^{-7}$ rad.

15·12 3×10^{-4} cm ; 5×10^{-4} cm ; $2{\cdot}5\times10^{-4}$ cm.

15·14 Claim doubtful, since angle subtended is less than θ_{min} of the eye (see Ex. 15.8).

15·15 Vertical separations not seen, horizontal ones clearly seen. (i) horizontal separations still seen, vertical not seen (iii) no separations seen.

15·16 $\sim10^5$, Fringes get too crowded to be viewed separately (problem of *geometrical* resolution).

15·17 $\sim$11/24 A°.

15·18 2·27 radian

15·19 20A°

16·3 61° ; 29°.

16·5 102°, 282° : $\frac{3}{4}$, $\frac{1}{4}$.

16·7 9·5. For -2° and $+2^\circ$ obsd. ratio 0·78 and 1·23 respectively.

16·12 $3{\cdot}2\,(n+\frac{1}{2})\mu$; $3{\cdot}2\,(n\pm\frac{1}{4})\mu\,(n=0, 1, 2\ldots\ldots)$.

16·13 As QW plate ; always (under the assumption stated).

16·15 QW for 7384, 5743, 4699 A°, HW for 5169 A°.

16·17 (i) *a* and *d*, *b* and *c*, none ; (ii) none, *b*, and *c* ; (iii) elliptical, plane-pol.

17·1 72·5° per dm per gm/cc, ; $\theta\approx22^\circ$.

17·3 +72°.

17·8 $\left|\frac{\Delta c}{c}\right|_{max}=\frac{\Delta\theta}{\theta}+\frac{\Delta\alpha}{\alpha}+\frac{\Delta l}{l}$. From the data $\Delta c/c\approx0{\cdot}01$ and $c=0{\cdot}284$. Range of error $\approx\pm{\cdot}003$ gm/cc.

17·9 transmission 3%, contrast 4%. For 5°, 0·76% and 8%.

17·10 $\alpha=49{\cdot}5^\circ$; $\theta=15{\cdot}8^\circ$.

18·3 $11{\cdot}14\times10^{-8}$ cm ; 0·076 mm/A.

18·5 50° 40′ ; 0·97 A° ; 12′.

18·7 1·8 A°, 7·2 cm^{-1}.

18·10 5·56 cm^{-1}, 14·8 cm^{-1} ; $\frac{\lambda}{\Delta\lambda}=\left|\frac{\nu}{\Delta\nu}\right|=3000, 1500.$

18·11 0·39 cm^{-1} ; 0·2 cm^{-1}.

18·12 4119·1 $A°$; 5889·1 $A°$

18·15 (i) 5×10^{-5} rad./$A°$ (ii) 16×10^{-5} rad/A and 30×10^{-5} rad/$A°$ for the first orders on two sides.

19.1 11 micron ; 8250 k.

19·3 energy $\sim7\times10^{-4}$ J ; quantum$\sim10^{-32}$ J.

19·4 $t=2\times10^{3}$ sec.

19·7 (i) $\nu>\sqrt{2}$ times, (ii) no change, (iii) v changes.

19·8 0·78 eV ; 0·28 micro amp.

19·9 $5{\cdot}43\times10^{-5}$ cm.

19·10 1·96 eV ; (i) no, (iii) no.

19·20 0·0363 A ; 0·054.

19·15 $9{\cdot}4\times10^{-7}$ Nm^{-2}.

19·16 $6{\cdot}78\times10^{-5}$ cm. $3{\cdot}43\times10^{-8}$ cm ; 3·64 volt

19·17 $1{\cdot}098\times10^{5}$ cm^{-1} ; 3940 A

19·18 14·0, 7·3 ; 5·1, 0·6 volt ; $1{\cdot}41\times10^{-8}$, $1{\cdot}87\times10^{-8}$ cm.

19·19 16·7 10^{-12} erg, $3{\cdot}92\times10^{-8}$ cm.

19·20 36000, 12000, 9000, 6000, 4000, 3600 $A°$

19·21 0·59 ; 0·34 ; 0·098.

20·1 $V=3{\cdot}3\times10^{6}$ volt
$E=3{\cdot}3$ Mev, 3·3 Mev, 6·6 Mev

20·2 ~780 layers ; $1{\cdot}5\times15^{15}$ atmos/cm^2 ; $\sim5\times10^{-3}$

20·6 In the region of interest $4<r\times10^{13}<8$ cm, one gets $V=0{\cdot}144$ [$(80/r)-12{\cdot}5$] Mev. Statement (i) alone fits the model.

20·7 $-0{\cdot}72$ eV

20·10 $1{\cdot}32\times10^{-8}$ cm

20·11 225$A°$; 300$A°$, 253$A°$

20·12 0·263A ; $13{\cdot}3\times10^{15}$ Hz : $1{\cdot}84\times10^{13}$ ms^{-1}

20·13 0 and 1 ; 1, 0, −1.

20·17 1542 $A°$

20·18 6·00, 6·60, 7·99V ; 941V.

20·20 $a_0/29$; $2a_0/11$

20·22 113 Kev ; 24 Kev ; 0·14$A°$

20·23 0·50$A°$

20·24 0·184A, 0·160A, 1·07A

21·2 (i) 6 cm, (ii) 0·05 $A°$. (iii) 10^{5}, (iv) 10^{3}

21·5 $\log f=-{\cdot}005$; $f\approx(1-{\cdot}005)$

21·7 $\sim 0{\cdot}8\%$

21·10 $1{\cdot}4\times10^{-6}$ cm^2 ; 1·7 *kw* /cm^2

22·1 $a=2{\cdot}57A^\circ$

22·4 $a^3/4$.

22·6 (i) $a=3{\cdot}36A^\circ$; (ii) 8 ; (iii) $a\sqrt{3}/2$.

22·8 $d_{100}=2{\cdot}82A^\circ$, $d_{110}=1{\cdot}99A^\circ$
$a'=b'=c'=a/\sqrt{2}$; $\alpha=90^\circ$, $\beta=60^\circ$, $\gamma=120^\circ$

22·11 orders 1, 2 and 3 from planes of $d=2{\cdot}94A^\circ$

22·12 2θ for $d_{100}=37^\circ18'$, $77^\circ34'$, $106^\circ42'$, $136^\circ0'$
for $d_{111}=68^\circ38'$, $119^\circ44'$; for $d_{211}=911^\circ0'$

22·15 relative $(\sin\theta)^{-1}=1 : 0{\cdot}71$, **0·50** : 0·41 : 0·355 ; Ignoring II orders (**underlined**), highest *d* values bear ratio 1 : 0·71 : 0·41, characteristic of bcc lattic.

22·17 $1{\cdot}672A^\circ$; $0{\cdot}167A^\circ$; $1{\cdot}282A^\circ$

22·20 Neutron diffraction, For ordered, (100) reflections of odd orders weaker than even by a factor f_6. For disordered all orders have same intensity. For partial order, intermediate state.

23·2 $\theta\pm4{\cdot}1\times10^{-33}$ rad

23·3 (a) (i) λ increases 10 fold, (ii) F_1F_2 increases 10 fold, (iii) ω unchanged. (b) (i) λ unchanged, (ii) F_1F_2 decreases 10 fold, (iii) ω increases 10 fold.

23·8 $a=27A^\circ$

24·3 $a=\dfrac{1}{2\gamma^2}={\cdot}005$; $\beta=0{\cdot}995$, $V=0{\cdot}4995c$

24·6 $L'=50{\cdot}0$ cm

24·7 1·00*m* × 1·00*m* no change)

24·11 $2{\cdot}990\times10^8 m/s$

24·12 1·0*m*×0·359*m*, length ∥ to motion appears shorter,

24·13 Zero, 0772, 0·993

24·14 $1{\cdot}07\times10^{-8}$s ; 3·11 m

24·17 0·549c, 0·979c, 0·9982c.

24·11 $1{\cdot}49\times10^{-18}$ kg. *m/s*

25·1 (a) $20\pi/3$, (b) $25\sqrt{3}/8$ lx, $E=15\sqrt{3}/8$ lx, $B=15\sqrt{3}/8\pi$ cd/m^2

25·2 (a) 250 π lm, (b) 2·5π, (c) 10/π cd/m^2

25·3 (a) 1600 cd, (b) ~71 lx, (c) ~38 lx

25·4 1 phot$=10^4$ lx ; 1 lambert$=10^4$ cd/m^2 ; $E=25$ m.cd

25·5 (b) $\phi_1=16\pi I/S'$; $\phi_2=16I/S'$

25·6 $d=4{\cdot}1$ cm, 5·2 cm, 16 cm, $E=960$ lx, 568 lx, 60 lx

25·8 58 cm from the weaker source between the sources
131 cm from the weaker source, outside the sources

25·9 4·4%

25·10 $3{\cdot}2\times10^{-6}$

25·11 5×10^9 lm/m^2, 9×10^7 lx

25·12 $\sim2\times10^6$ cones/cm^2.

Index

A

Abbe's sine condition 56, 75
Abbe's theory of resolution 236
Abberrations, chromatic 37, 77
—lateral 39
—longitudinal 37
Aberrations, monochromatic 45
—astigmatic 46, 57
—coma 46, 56, 74
– curvature of field 47, 58
—distortions 47, 59
—of thin lens 50
—reduction of, 55, 57, 58
—speherical 45, 48, 131
Absorption 288, 323, 343
—inverted 343
Achromatism of Lenses 68, 70
— of size, not position 42
Alpha scattering 316
Angular momentum, orbital 321 410
—operator for 405
—spin 327
Anisotropy, optical 97, 248
Annihilation of mass 431
Aplanatic points 10, 51, 55
— lenses 51, 55, 75
—surfaces 51, 54
Aspherical lenses 56
Astigmatism 46, 57, 67
Atom, Bohr model of 321
—electronic config, 329, 331
—model for general 326
—Rutherford model 320, 325
—shell model 326,
—Thomson model 316
Atomic spectra 287, 293
—series relations in 303
Atomic structure factor 374

B

Balmer series 305, 326
Barrier penetration 408
Basis 367
Biprisn, Fresnel's 146
—expt. with electrons 389
Biquartz device 275
Blackbody radiation 293, 294
Bohr's theory 320, 410
—and de Broglie eqn. 391
Bohr radius 322
Boltzmann Law 296, 324, 343
Born's interpretation 393, 402
Bragg's Law 376
Bragg spectrometer 377
Bravais lattices 365
Brightness of images 435, 437, 439
Brewster's Law 245

C

Candela 434, 435
Candle power 438
Cardinal points 13, 17 68, 70
– deviation method for 33
– of eyepieces 68, 70
—of lens combination 25, 32
Cavity resonator 365
Cerenkov radiation 132
Characteristic impedance 96
Chromatic aberration (see aberration)
Coherence, spacial 341
—temporal 340
Coherence length 340
—time 340
[Coherent illumination 236

Coherent sources 140, 340
Collisions, inelastic 308
Colours of thin films 157
Coma 46, 56, 74
Complex amplitude 107
    —frequency 109
    —representation 116
Complementarity 393
Compton effect 300, 335
    —wavelength 302
Concave grating 212
    —mountings of 213, 215
Conjugate relationships 15, 18
Convention of signs 13
Coordination number 367
Core electrons 330
Cornu's prism 283
Cornu's spiral 190
Corpuscular theory 90
Cotton-Mouton affect 276
Coulomb's Law, 93
    —in alpha scattering 319
Crystal structure 363, 367
    —basis 367
    —CsCl 368
    —diamond, ZnS 370
    —hcp, NaCl 369
Crystal structure factor 375
Crystals, positive and negative 250
    —seven systems of 364
Curvature of the field 47, 58, 74

D

Davisson & Germer's expt. 381
De Broglie waves 236, 380, 386, 397
    —and Bohr condition 391
    —and particle in a box 391
    —Born's interpret. of 392
Diffraction (of light) 91, 176, 193
    —at a circular aperture 179, 200
    —at a single slit 194, 399
    ——angular spread in 198
    ——method of calculus 197
    ——phaser method 195
    —at N slits 203
    —at a straight edge 185
    —at two slits 202
    —Fresnel and Fraunhofer 193
    —of X-rays 375, 383
    —ripple tank expt. 176
Diffraction of electrons 380, 383, 390
    —Taylor's expt. 394
Diffraction of *n*, *p*, He 380, 381
Diffraction grating 207
    —reflection type 210
    ——blazed 210
    —resolving power of 225, 240
    —spectra, special features of 211
Dipole radiation 98, 101
Displacement current 93
Dispersion 36, 208, 268
Dispersive powers, table of 37
Distortion in images 47, 59
Division of amplitude 153
Double image prisms 257
Double refraction 248, 253
    —electromagnetic theory 251
    —Huygens theory of 249
Duality (wave-particle) 386, 403
Duane and Hunt limit 336
Duker's experiment 390
Dulong & Petit's Law 311

E

Eigenfunctions 403
Eigenvalues 403
Einstein's A & B Coeffs, 343
Einstein's theory of photoelectric effect 297
    —of relativity 414
    —of specific heat of solids 311
Electromagnetic theory 90, 93
Electromagnetic waves 90
    —different regions of 279
    —emission of 97
    —energy density in 100

—transverse nature of 95
Electronic configuration 329, 331
Electron microscope 76, 236
Electrons, refraction of 388
Electron spin 327
Emission 287, 303, 321
- stimulated 342
Emissivity, spectral 294
Energy density in waves 100
—spectral 294
Energy level diagram 306, 310, 323, 325, 333, 334
—of H atom 323
—cf Hg, Cs, Na, H 310
—of particle in a box 391
—of Ruby 348
Equivalence of m & E 430
Ether, lumineferous 91
—motion of earth through 169, 416
—transverse waves in 91
Excitation potentials 324
Extraordinary rays 249
—refractive index 251
Extremum path 7
Eye, the Human 441, 445
Eye lens 66
Eyepieces 65
—achromatism of 68, 70
—Gaussian 71
—Huygens' 66, 74
—Kellner's 72, 74
—need of multiple lens 65
—Ramsden's 69, 74
Expectation value 398, 405
Exponent representation 112, 341

F

Fabry & Perot, interferometer 172
—etalon 172, 228
Faraday effect 276
Fermat's principle 1, 2, 6, 63
—and aplanatic points 10
Field lens 66
Field of view 62, 69, 71
Fluorescence 291
Focal length 16, 21, 26, 32
Franck & Hertz expt. 307, 311, 318
Fourier series 110, 113, 200, 231
—exponential repres, of 111
Fourier transform 112, 341
—spectroscopy 230
Fraunhofer diffraction 193
(see also diffraction)
Fresnel's hp zones 178, 185
Fringes 143
—achromatic 149
—in parallel films 163
—in wedge film 156
—lateral displ. of 148
—of equal inclination 163
—of equal thickness 155
—zero order 168

G

Galilean transformation 414
Gauss's Theory 14
Gaussian telescope 31
Geometrical optics 1
Grating (see also diffraction)
—at oblique incidence 284
—blazed 210
—concave 212, 215, 284
—dispersion in 208
—overlapping orders in 211
—plane 207, 211
—resolving power of 212
—wire 220
Group velocity 120, 407

H

Haidinger's fringes 163
Half-period (hp) zones 178
—on cylindrical w.f. 185
Half-shade device 274
Half-wave plate 258
Hamiltonian operator 405
Harmonic generation 358
Harmonic oscillator 105, 107, 407
—Complex amplitude 106

—in quantum mechanics 407
Heisenberg's principle 396
—γ ray microscope 399
Helium, diffraction of 381
—spectrum of ionized 324
Helmholtz's law 20, 24
Holography 354
Huygens' principle 124, 250
—limitations of 125, 177
Hydrogen atom, 320, 391
—radial distbn. fn. 410
—spectral series of 305

I

Illuminance 435, 439
Induced emission 342
Infrared spectrographs 284
Intensity distribution
(see Interference and Diffraction)
Intensity in waves 100
Interference 136, 153
—in thin films 154, 157
——colours due to 157
—in space 139
—in white light 147
—of two sources 137, 142
—of several waves 170
—Newton's rings 159
—Young's expt. on 144
——an equivalent to 145
Interferometer, F-P 172
—Michelson 165
—Twyman & Green 169
Interferometry, multiple beam 170
Ion microscope 236
Ionization potential 324

K

Kerr cell method 86
Kerr effect 277
Kinetic energy, relativistic 428

L

Lagrange's law 20
Lambert's law 436
Lasers 339, 348
—coherence, of 341
—directionality of 351
—dye/semiconductor 349
—gas/glass 349
—monochromaticity 350
—ruby 347
—shaping of beam in 345
—uses of 352
Lateral displacement 148
Lattice, Bravais 365
—planes 372, 375
——spacings of 375
Laue procedure 375, 378
Length contraction 421
Lens, bending of 55
—combination 28, 30, 32
——achromatic 40
——telephoto 30
——telescopic 30
—meniscus 27, 54
—shape factor of 50, 55
—symm. doublet 60
—thick 25,
—thin 19
Light, corpuscular theory 90
—e.m. theory 90, 93
—laws concerning 1
—quantum theory 103, 306, 321
—ether theory, 91
Llyod's mirror expt. 147
Lorentz transformation 97, 418
Lumen 434
Luminescence spectra 291
Luminous flux 434
Luminous Intensity 434
Luminous sensitivity 440
Lummer & Gehrcke plate 230

Lux 434
Lyman series 305, 323

M

Magnification 20, 21
 —and resolution 222
 —lateral 21
 —normal 75, 232, 235
Malus, law of 245, 275
Mass, variation of 426
 —energy equivalence 430
Matrix mechanics 294, 403
Matter waves (see De Broglie)
Maxwell's equation 93
 —invariance of 97, 419
Meniscus lens 27, 54
Michelson's echelon 231
Michelson's interferometer 165, 233, 340, 351
 —comparison with F.P. 172
Michelson's method for c 85
Michelson-Morley expts. 169, 416
Microscope, compound 75, 77
 —coherent illumination 236
 —dark field illumination 76
 —electron/ion 76, 236
 —high power objective 75
 —numerical aperture 75
 —oil immersion 11, 25, 53, 236
 —phase contrast 237
 —resolution in 234, 236
 —ultra 76
 —γ-ray 398
Miller indices 372
Millikan's expt. 299
Monochromators 279, 281
Moseley's law 334
Multiple beam int. 170
Multiple reflection 153

N

Nearest neighbour 367
Negative temperature 344
Neutron diffraction 382, 383
Newton's formula 16
Newton's rings 91, 159
 —in transmitted light 162
Nicol's prism 255
Nodal points 17
Nodal/antinodal planes 139
Non-linear Optics 357
 —spectroscopy 359
Non-proper measurements 421
Non-reflective coatings 157
Normalization of ψ 407
Nuclear size 318
Numerical aperture 75, 235

O

Oil immersion objectives 11, 23, 53, 236
Operational view 420
Operators in QM 401
 —table of 403
Optic axis 250, 252
Optical activity 266
 —Fresnel's expln. of 270
 —origin of 268
 —Table of 268
Optical anisotropy 252
Optical instruments 62
 —resolving power of 221
Optical path 3, 131
Optical pumping 344
Optical rotation (see Optical activity)
Optical systems 13
Ordinary ray 249
Overlapping spectra 211

P

Packing fraction 371
Pair production 431
Paraxial rays 45
Particle in a box 391, 399, 407

Pauli's principle 329, 357
Penetrating orbits 332
Periodic Table 330
Petzval combination 58, 59
Phase 116, 120
  —change at reflection 133, 158
  —randomness of 118
Phase contrast microscope 237
Phase retardation plates 257
Phase velocity 117, 120, 387
Phase waves (see De Broglie waves)
Phaser diagram method 138, 190, 195
Phosphorescence 291
Phot 439
Photoconductive cell 300
Photoelectric cell 297
Photoelectric effect 293, 297, 325
  —Einstein's theory 297
  —Millikan's expt. 299
Photometers 442
  —Bunssen's 442
  —Lumer—Brathan 443
  —Flicker 444
  —Ulbright's 445
Photometry 434
Photomultiplier 300
Photon concept 298
Photovoltaic cell 300
Pile of plates 247
Planck's theory 294
Polarimeters 273
Polarization 96, 119, 245
  —analysis of 261
  —by double refraction 248
  —by reflection 245
  —by transmission 247
  —different kinds of 250, 265
  ——production of 260
  —rotatory (see opt. activity)
Polorizer' wire grid 243
Polaroids 254
Population inversion 344, 349
Powder method (X-rays) 378
Primitive cell 365
  —vectors 366
Principal planes/points 16
Principal section 256
Prism, constant deviation 281
  —Cornu's 283
  —double image 257
  —materials 284
  —Nicol 255
  —resolving power of 226
Probability interpretation 393, 402
  —waves (see De Bronglie)
Proper measurements 421
Pulse 114, 394
Pumping, optical 344
Pupil, entrance & exit 62, 66, 77

Q

Quality factor 108, 118, 340
Quantum defect 331
Quantum Mechanics 103, 293, 326, 386, 403
  —expectation value in 398, 405
  —of H atom 408
  —of harmonic oscillator 407
  —of particle in a box 406
  —postulatory basis of 404
Quantum numbers n, 1, m, 327
Quarter wave plate 258
Quartz spectrograph 282

R

Radar, resolving limit of 283
Radiation by atoms 117
  —blackbody 294
  —due to acc, charge 98
  —due to dipole osc. 98
  —due to osc, charge 98
Raether's expt. 390
Raman spectroscopy 289, 353, 360
Rational & Irr. spectra 211
Rayleigh criterion 223
Rayleigh-Jeans law 295, 297
Rayleigh refractometer 149
Rays, ord, & extraord, 248
  —mutual indep. of 106
  —reversibility of 1

Recoil electron 300
Rectilinear prop. 1, 90, 184
Reflection, 4
  —at a sph. surface 18
  —and Fermat's law 5
  —Fresnel's 132, 153
  —multiple 153
  —phase change at 133, 153
Reflection of waves 127
Refraction, and Fermat's law 6
  —double 248, 251, 254
Refraction of waves 128
Refraction of electrons 388
Refractive index, O & E 250, 254
Relativity 414, 429
  —and length contraction 421
  —force and KE in 429
  —mass variation in 424
  —operational view of 4[illegible]7
  —time dilation in 421, 424
  —velocity addition in 426
Relaxation time 108
Resolving power/resolution 212, 221
  —and magnification 222
  —geomet. & spectral 221
  —of eye 232
  —cf F.P etalon 228
  —of grating 225, 239
  —of LG plate 230
  —of microscope 75, 234
  ——Abbe's theory of 236
  —of prism 226, 239
  —Rayleigh's criterion for 223
Resonance half-width 109
Rest mass 427
Retardation plates 257
Retarded time 97
Ripple tank 134, 176
Ritz combination principle 304
Rotation, specific 267
  (see also optical activity)
Rowland grating
  (see concave grating)
Rutherford scattering 316
Rydberg constant 304, 324

S

Sachharimeter 273
Schmidt correction plate 56
Schrodinger's equation 400
  —applications of 405
  —for stationary states 403
Secondary wavelets 125
Shape factor 50
Shell model 316, 326
  —and periodic table 330
Shock waves 132
Signs, convention of 13
Sinc function 113, 198
Snell's law 7, 129
Sonar 233
Space lattice 363, 364
Specific heat of solids 293, 311
Specific rotation 267
  —Table of 258
Spectra 279
  —absorption 286
  —emission 287, 306
  —luminescence 291
  —of alkali atoms 331
  —of He$^+$ 324
  —Raman 289
  —Rational & irr. 211
  —series relations in 304
  —X-ray 333, 336
Spectral lines, purity of 339, 351
Spectral resolving power
  (see resolving power)
Spectrographs 279
  —constant deviation 281
  —for infrated 284
  —for ultraviolet 283
  —Littrow 280
  —quartz 282
  —vacuum 283

- —X-ray 377
- Spectrophotometer 287
- Spectroscopy (see spectra)
- Speed of light, methods for 80-88
- Spherical aberration (see Abberation)
- Spin-orbit coupling 328
- Spontaneous emission 342
- Standardization of metre 169
- Stellar Interferometer 233
- Stimulated emission 342
- Stokes treatment 134
- Stopping potential 299
- Stops 57
- Superposition principle 136

T

- Taylor's expt. 394
- Telephoto lens 30
- Telescope 30, 76
  - —Gaussian 31
  - —magnification in 77
  - —resolving power 231, 240
  - —Stars seeing through 441
- Thin films, colours of 154
- Third order theory 50
- Thomson's expt. 382
- Time dilation 421
- Tourmaline plate 243
- Translational group 363
- Tunneling 408
- Twymann-Grren Int. 169

U

- Ultramicroscope 76
- Ultraviolet catastrophe 295
- Uncertainty principle 395
  - —and diffraction at a slit 399
  - —and particle in a box 401
  - —in Bohr orbits 401
- Uniaxial crystal 252
- Unit cell/vector 363, 364

V

- Valence electrons 330
- Velocity of light (see speed)
- Velocity, group and wave 120
- Velocity, relativistic transformation, 425
- Verdet's constant 276
- Vibrations and waves 105
- Visibility in fringes 162

W

- Wave equation 115, 400
- Wave function 400
  - —acceptability 404
  - —meaning of 393, 402
  - —normalization 405
  - —time development of 407
- Wave front 116
  - —in uniaxial crystals, 251
- Wave group 118, 396
- Wave mechanics (see QM)
- Wave normal 252
- Wave number, 289
- Wave packet 118
  - —Fourier transform of 112
- Wave-particle quality, 386, 394
- Waves 115
  - —from atoms and molecules 117
- Waves, reflection and refraction 126
- Wave trains 120
- Wave velocity 120, 122
- Wave theory 105, 125, 130, 137
- Wein's displacement law 295
- Wire gratings 210
- Wire grid polarizer 242
- Work function 297

X

X-ray, Bragg spectrometer 377
  —diffraction 375
    ——structure factor in 375
  —Laue method 370
  —powder method 378
X-rays, origin of 333
  —Compton scattering of 300
X-ray spectra 333
  —absorption 335
  —characteristic 333
  —continuous 336
  —Duane and Hunt limit 336
  —determination of λ 376
  —Moseley's Law 334

Y

Young's experiment 144
  —simple equivalent of 145

Z

Zero-point energy 407
Zone plate 182
  —multiple focii in 184
  —phase reversal 184

[illegible]
8-6

OUR OUTSTANDING PUBLICATIONS

Sharma & Sankar
Thermodynamics and Statistical Physics

Tayal. D.C.
Electricity and Electronics

Tayal, D.C.
Nuclear Physics

Tayal, D.C.
Electronics

Bhattacharya. P.K.
Group Theory and Its Chemical Applications

Chatwal & Anand
Instrumental Methods of Chemical Analysis

Chatwal, G.R.
Synthetic Organic Chemistry

Chatwal, G.R.
Synthetic Dyes

Chatwal. G.R.
Synthetic Drugs

Chatwal & Anand
Spectroscopy (Atomic and Molecular)

Chatwal. Gurdeep
Organic Chemistry of Natural Products Vol. I, II

Chatwal, G.
Reaction Mechanism and Reagents in organic Chemistry

Arora, M.P.
Animal Physiology

Arora, M.P.
Genetics

Goswami. H.K.
Practical Cytology – Applied Genetics and Biostatistics

Goel, S.C.
Principles of Animal Developmental Biology

Powar, B.C.
Cell Biology